

**A multi-scale exploration of the relationship
between spatial network configuration and
housing prices using the hedonic price
approach.**

A Greater London case study

Stephen Law

The Bartlett School of Architecture

UCL

Submitted for the degree of Doctor of Philosophy

September 2017

Copyrights

I, Wai Pan Stephen Law confirm that the work presented in this thesis is my own.
Where information has been derived from other sources, I confirm that this has
been indicated in the thesis.

Signature.....

Copyright © Stephen Law 2017

All the Ordnance Survey map data that has been used and reproduced in this thesis are covered under the following copyright information: © Crown Copyright and database right 2017. An Ordnance Survey/EDINA supplied service.

All the Census data that has been used and reproduced in this thesis are covered under the following copyright information: National Statistics data © Crown copyright and database right 2017

All the house price data that has been used and reproduced in this thesis are covered under the following copyright information: Land Registry © Crown copyright and database 2017

All the land use data that has been used and reproduced in this thesis are covered under the following copyright information: Valuation Office Agency under the Open Government License v1.0 © Crown copyright 2017

All the Nationwide transaction data was provided by Nationwide through a licensing agreement with London School of Economics (LSE). © Nationwide Building Society

I dedicate this thesis to Kelin, my fiancé
for your unconditional love,
for your patience and
for your intelligence.

Acknowledgements

I am whole-heartedly thankful to my two supervisors, Professor Alan Penn and Dr Kayvan Karimi, for their encouragement, their guidance and their supervision which enabled me to complete my thesis. I am also thankful to Professor Laura Vaughan, Professor Bill Hillier and the many colleagues and friends at the UCL Bartlett Space Syntax Laboratory, past and present, whom motivated me and inspired me.

I would also like to give my special thanks to Space Syntax Limited and the many colleagues and friends there, past and present: Tim, Anna, Max, Christian, Nora, Antonin, Ji, Fefe, Ed, Godfrey, Laurens, Ioanna, Abhi, Raman, Kaveh and Su. Thank you all for your patience and the intellectual capacity to pursue both a professional and academic career simultaneously.

I sincerely thank Professor Stephen Gibbons at LSE for his advices on the econometrics, to Dr. Jorge Gil for your encouragement and advice specially on chapter 6, to Professor Lars Marcus for his interest in my work, to Dr. Alain Chiaradia for giving me guidance throughout the thesis, to Dr. Yao Shen for walking this PhD journey together with me and to the many others whom have helped me a great deal in the past five years.

Finally, I would like to thank my father, my mother, my sisters whom I am incredibly grateful for their support and their unconditional love throughout this process.

Abstract

Using the hedonic price approach, the prices of heterogeneous goods such as housing can be derived from the sum of the item's utility-bearing parts. Since its introduction, the approach has become an established real estate method for valuing intangible goods. Despite the well-known impacts of amenity and access on house prices, limited attention has been given to the impacts of the urban built form on property values. This stems partly from the fact that few quantitative methods exist that can provide a thorough understanding of the built environment. Through three interrelated strands of empirical research, this research proposes using spatial configuration methods and the hedonic price approach as an empirical strategy for enhancing the analysis and interpretation of the housing market in the densely populated region of Greater London.

The first strand of the argument focuses on the correlation between accessibility and house prices. Existing results show a strong relationship between geographic accessibility, such as Distance to the Central Business District and house prices. There are, however, two problems with this approach. First, geographical measures assume a predetermined employment location; second, is that such measures often fail to consider the network effects on house prices. This study reveals that spatial network centrality and geographical measures are jointly significant in explaining house prices. The second strand of the argument focuses on the relationship between the local area and house prices. In the past, census output areas were used for measuring this relationship. In reality, however, utilising these arbitrary definitions has led to inconsistent results. In contrast, the proposed study takes a Street-based Local Area (St-LA)" approach, which can more accurately captures the subtle differences in the urban environment. In response to the common problem of using census output areas as building blocks for defining housing submarkets, the third strand of this research shows that housing submarkets can be described more accurately using the St-LA approach through the application of a hedonic price model. The results demonstrate that spatial configuration factors are significant when correlated with house prices at different scales.

This research links the results of the three areas of investigation in order to provide a more comprehensive understanding of the economic performance of the built form and to encourage planners, decision-makers and developers in using spatial configuration-based methods in planning and design that can lead to more equitable and sustainable policy making.

Table of Contents

Chapter 1 Introduction.....	18
1.1 Background	18
1.1.1 Urban Rent Theory.....	19
1.1.2 Space Syntax Theory.....	20
1.1.3 Hedonic Price Model as the Empirical Approach.....	21
1.1.4 Greater London Case Study	22
1.2 Empirical Strategy	22
1.2.1 Accessibility.....	23
1.2.2 Local Area	24
1.2.3 Housing Submarkets.....	24
1.3 Thesis Outline	25
1.3.1 Summary.....	27
Chapter 2 Review of the Hedonic Price Model.....	28
2.1 Background	28
2.2 Key Theoretical Concepts	30
2.2.1 Spatial Equilibrium and the Monocentric Bid Rent Model.....	30
2.2.2 Hedonic Price Theory and Intra-city House Price Model	31
2.2.3 The Housing Submarket Theory	32
2.2.4 The Rosen-Roback Inter-regional Model	32
2.2.5 The Limitations of the Hedonic Price Model	33
2.3 The Empirical Approach of the Hedonic Price Model	34
2.3.1 The Dependent Variable: House Prices.....	35
2.3.2 The Independent Variable Set	36
2.3.3 Structural variable	37
2.3.3.1 Structural variable limitation	37
2.3.4 Location Variables.....	38
2.3.4.1 Access to Employment	38
2.3.4.2 Access to Public Transport	38
2.3.4.3 Access to Private Transport.....	39
2.3.4.4 Location variable limitation.....	39
2.3.5 The Neighbourhood Variable	40
2.3.5.1 Environmental Amenities	40
2.3.5.2 Urban Amenities.....	41
2.3.5.3 Neighbourhood Effect	42
2.3.5.4 Neighbourhood variable limitation.....	42
2.3.6 The Housing Submarket	43
2.3.6.1 Hierarchical nature of housing submarkets	43
2.3.6.2 Local Area Definition	44

2.3.6.3 Submarket Identification	44
2.3.6.4 Limitations in housing submarket research	45
2.3.7 Other Controlled Variables	46
2.3.8 Inter-city Regional House Price Variables	46
2.3.8.1 Wage, Productivity and Population	46
2.3.8.2 Supply-side, Physical Constraint and Construction Costs	47
2.3.8.3 Policy and Investment Factors	47
2.3.9 Limitations in empirical studies using the hedonic price approach	48
2.4 Empirical Specification	48
2.4.1 Multiple Variable Regression Model	49
2.4.2 Regression Model Limitations	49
2.4.2.1 Multicollinearity	49
2.4.2.2 Heteroscedasticity	50
2.4.2.3 Omitted Variable Bias	50
2.4.2.4 Spatial Autocorrelation	50
2.4.3 Advanced Regression Models	51
2.4.4 Spatial Regression Models	51
2.4.5 Panel Data Fixed Effect Regression Models	52
2.4.6 Multilevel Regression Models	52
2.5 Discussion	52
2.6 Research Gaps - Spatial Configuration	53
Chapter 3 Spatial Configuration Methods	55
3.1 Introduction	55
3.1.1 Space Syntax	55
3.1.2 Space Syntax as a Theory for Cities	55
3.1.3 Three Research Strands in Linking Spatial Configuration and House Prices	56
3.1.4 Discussion	58
3.2 Research Strand One: Geometric Accessibility	59
3.2.1 Spatial Network Model	60
3.2.2 Spatial Dual Graph Representation	61
3.2.3 Accessibility Definition	62
3.2.3.1 Geographic and Geometric Accessibility Methods	62
3.2.3.2 Distance and Radius	64
3.2.4 Discussion and Limitations	65
3.3 Research Strand Two: Street-based Local Areas (St-LAs)	65
3.3.1 Strand Two Method	66
3.3.2 UK Statistical and Administrative Geography	67
3.3.2.1 UK Statistical and Administrative Local Area	68
3.3.2.2 Postcode Unit	69
3.3.2.3 Super Output Area (LSOA/MSOA)	69

3.3.2.4 Electoral and Census Area Statistic Ward	70
3.3.3 Street-based Local Areas	70
3.3.3.1 Definition	70
3.3.3.2 Previous Syntactical Methods	71
3.3.3.4 Community Detection Techniques	72
3.3.3.5 The Modularity Optimisation Algorithm	73
3.3.3.6 The Walktrap Algorithm	74
3.3.3.7 Infomap Algorithm	75
3.3.3.8 The Spin Glass Algorithm	76
3.3.3.9 Community Detection Methods Comparison	76
3.3.4 Discussion	79
3.4 Research Strand Three: Street-based Housing Submarkets (St-HS)	80
3.4.1 Strand Three Method	80
3.4.2 The Housing Submarket Definition	81
3.4.3 Statistical Clustering and Algorithms	81
3.4.3.1 The K-means Clustering Algorithm	82
3.4.4 Limitations	82
3.5 Summary	83
Chapter 4 The Research Framework, the Case Study and the Datasets	84
4.1 Background	84
4.2 The Hedonic Price Regression Model	84
4.2.1 Multi-level Analytical Framework	84
4.2.2 Software	85
4.3 Case Study.....	85
4.3.1 Location.....	85
4.3.2 London's Government.....	86
4.3.3 London's Demographics	87
4.3.4 London's Economy.....	87
4.3.5 London's Transport	88
4.3.6 Discussion	90
4.4 The London Housing Market.....	90
4.4.1 London House Prices.....	91
4.4.2 The London Housing Crisis.....	92
4.4.3 The Future of London's Housing.....	96
4.4.4 Discussion	97
4.5 Variables and Dataset.....	97
4.5.1 Research Strand One Dataset.....	99
4.5.1.1 The House Price Dataset.....	99
4.5.1.2 The London Street Network and the Transport for London Network Datasets	104
4.5.1.3 The London Retail Address Dataset	106

4.5.1.4 The UK School Performance Dataset.....	108
4.5.1.5 The London Heritage Parks and Gardens Dataset.....	110
4.5.1.6 UK Census Employment Dataset	111
4.5.2 Research Strand Two Datasets	111
4.5.2.1 The Spatial Street Network Dataset.....	112
4.5.2.2 Postcode Unit Dataset	113
4.5.2.2 Census Area Statistic (CAS) Ward	114
4.5.2.3 Super Output Area (LSOA, MSOA)	115
4.5.3 Research Strand Three Dataset	116
4.5.4 Discussion	116
Chapter 5 Spatial Network Accessibility Effects on House Prices	118
5.1 Introduction	118
5.1.1 Strand One Research Question Definition.....	119
5.2 Geometric Accessibility Framework	120
5.2.1 Accessibility Specification	120
5.3 Empirical Method	123
5.3.1 Fixed-Effect Hedonic Price Approach	123
5.4 Dataset and Study Area	127
5.4.1 Greater London Area	127
5.4.2 House Prices Dataset	128
5.4.3 Transport Innovation Projects Between 1995 and 2011	131
5.4.4 Descriptive Statistics	133
5.5 Empirical Results	134
5.5.1 The OLS Regression Model Results.....	134
5.5.2 Fixed-Effect Regression Results.....	136
5.5.3 Fixed-Effect Regression Model Results for Testing Joint Accessibility Effects	137
5.5.4 Multi-collinearity tests between geometric and geographic accessibility measures	138
5.6 Discussion	139
5.6.1 Limitations	140
5.6.2 Conclusion	141
Chapter 6 Street-based Local Area Effects on House Prices	142
6.1 Introduction	142
6.1.1 Strand Two Research Question Definition.....	142
6.2 Framework for Street-based Local Area	144
6.2.2 Community Detection Methods	144
6.2.3 Defining St-LA Using the Street-network Dual Graph.....	145
6.2.4 Modularity Optimisation Algorithm on the Street-network Dual Graph	145
6.2.5 Street-based Local Area Subgraph Network	146
6.3.1 Intra-cluster House Price Analysis	148
6.3.2 Inter-cluster House Price Analysis	148

6.3.3 Multilevel Hedonic Price Approach	149
6.3.4 Model Comparison: Street-based Local Area and Region-based Administrative Local Area.....	154
6.4 Datasets and Study Area	154
6.4.1 Greater London Area	154
6.4.2 Residential Sold Price	155
6.4.3 London Street Network	155
6.4.4 London Street-based Local Area	156
6.4.5 London Administrative Local Area Units	157
6.4.6 Descriptive Statistics	158
6.5 Empirical Results	158
6.5.1 Intra-cluster House Price Analysis Results	159
6.5.2 Inter-cluster House Price Analysis Results	160
6.5.3 Street-based Local Area Multilevel Regression Results	163
6.5.4 Comparison of Street-based Local Areas and Administrative Local Areas	166
6.6 Discussion	167
6.6.1 Benefits and Limitations	169
6.6.2 Conclusion	171
Chapter 7 Street-based Housing Submarket.....	172
7.1 Introduction	172
7.1.1 Housing Submarkets	172
7.1.2 Strand Three Research Question Definition	173
7.2 Street-Based Housing Submarket Framework	174
7.2.1 Submarket Construction Procedures	174
7.2.2 Unsupervised K-means Clustering Algorithm	175
7.3 Empirical Method	176
7.3.1 Housing Submarket Test.....	176
7.3.2 Multilevel Hedonic Price Regression Model	177
7.3.3 Comparing Street-based and Traditional Housing Submarkets	181
7.4 Dataset and Study Area	182
7.4.1 Greater London Area	182
7.4.2 London Local Area Units.....	183
7.4.3 Descriptive Statistics	183
7.5 Empirical Results	184
7.5.1 The Housing Submarket Tests	184
7.5.2 London Street-Based Housing Submarkets (St-HS)	185
7.5.3 London St-LA Submarkets Individual Regression Results	187
7.5.4 Multilevel Hedonic Price Model Results	188
7.5.5 Street-based Housing Submarket Model Comparison	190
7.5.6 Street-based Housing Submarket and Traditional Housing Submarket Model Comparison	191
7.6 Discussion	193

7.6.1 Conclusion and Limitations	194
Chapter 8 Conclusion	196
8.1 Introduction	196
8.1.1 Chapter summary.....	196
8.2 Key Findings	196
8.2.1 Spatial Network Accessibility	197
8.2.2 Street-Based Local Areas	199
8.2.3 Street-Based Housing Submarket	201
8.2.4 Summary and Potential Causal Inferences	204
8.3 Research Limitation	204
8.3.1 Research Strand One Limitations	205
8.3.2 Research Strand Two Limitations	206
8.3.3 Research Strand Three Limitations	207
8.4 Future Research Directions	207
8.4.1 Case Study Across Geographies.....	207
8.4.2 Fuzzy Spatial Network Local Area Boundaries.....	209
8.4.3 Simulating House Price Spillover Effects.....	210
8.4.4 Estimating the Economic Value of the Urban Design	211
8.5 Research Implications.....	212
8.5.1 Implications for Planners.....	212
8.5.2 Implications for Developers.....	213
8.5.3 Implications for Space Syntax.....	213
8.6 Conclusions.....	214
Appendix A Charts made by the author based on data from Sirman et al. (2006).	216
Appendix B Parks and gardens in London (English Heritage 2014).	217
Appendix C Global Moran's I	218
Appendix D Implicit prices of accessibility for individual cities in the UK (Law et al. 2017a).....	219
Bibliography	220

List of figures

Figure 1.1 The limits of zonal representation of space.....	20
Figure 1.2 Two configurations of the same house conceptualised in a graph (Hillier and Vaughan, 2007).	21
Figure 1.3 Intra-city hedonic price model. Spatial configuration effects highlighted in red.	23
Figure 2.1 Map of the review of the hedonic price model.	29
Figure 2.2 Key theoretical concepts for the hedonic price model.....	30
Figure 2.3 The bid rent curve. Credit: by SuzanneKn at Wikipedia (Wikimedia Commons).....	31
Figure 2.4 The hedonic price empirical approach.....	35
Figure 2.5 Intra-city hedonic price independent variables.	36
Figure 2.6 Empirical specification for the hedonic price regression model.....	49
Figure 2.7 The review identified three research gaps concerning the association between spatial configuration factors and house prices.	53
Figure 3.1 Conceptualising space syntax as a theory for city. Diagram has been adapted from Wegenaar's (1994) LUTI diagram.	56
Figure 3.2 The three research strands.	59
Figure 3.3 Geometric accessibility considers the network effect and geographic accessibility considers attraction and distance effect.	60
Figure 3.4 Accessibility analysis explanation for the first research strand.....	60
Figure 3.5 The spatial network graph definition.....	61
Figure 3.6 On the left is the traditional administrative local area, which does not consider the network attribute of the street network (red) and on the right is the Street-based-Local-Area (green).	66
Figure 3.7 The St-LA unit and the traditional region-based local area unit.....	67
Figure 3.8 The UK administrative, census and postal geography classes (ONS 2015).....	68
Figure 3.9 Postcode unit, LSOA, MSOA and ward visualised in Thamesmead	69
Figure 3.10 The modularity optimisation algorithm. (diagram produced by the author).	74
Figure 3.11 A map of the social sciences (Rosvall and Bergstrom 2007).....	75
Figure 3.12 The study area boundary.	77
Figure 3.13 A visualisation matrix between the Walktrap, modularity optimisation, Spin glass and Infomap algorithms for the four named urban areas in London.....	78
Figure 3.14 St Anne's Church, Soho (London County Council 1966).	79
Figure 3.15 On the left is the St-HS that combines street network attributes, spatial attributes and structural attributes; on the right are traditional housing submarkets formed by spatial-structural attributes.	80
Figure 3.16 The submarket construction process.....	81
Figure 4.1 The hierarchical nature of the housing market.	85
Figure 4.2 The Greater London boundary.....	86
Figure 4.3 The 32 London boroughs and the City of London.....	86
Figure 4.4 London population between 1801 and 2011.....	87
Figure 4.5 London industry sectors in 2011.....	88
Figure 4.6 The London Underground, the National Railway system and the street network. Author produced the diagram.	89
Figure 4.7 London transport modes	89
Figure 4.8 Dwelling type in Greater London.	90
Figure 4.9 Number of bedrooms in Greater London.	91

Figure 4.10 Dwelling age in Greater London.	91
Figure 4.11 London house price and transaction data	92
Figure 4.12 London Housing Crisis' appearance on Google search forms from 2010 to 2015. Source: Google, 2015	93
Figure 4.13 Greater London's population between 1801 and 2011.....	93
Figure 4.14 Greater London Population and Employment Distribution.....	94
Figure 4.15 The percentage change in the number of jobs, population and homes from 2009 to 2014.	95
Figure 4.16 The median house price to earnings ratio in Greater London. Source: Department for Communities and Local Government	95
Figure 4.17 London opportunity areas and areas for intensification.Source: GLA (2015).	96
Figure 4.18 House price distribution in 2011.....	101
Figure 4.19 Visualisation of London House Price in 2011, from red indicating high to blue indicating low.....	102
Figure 4.20 Visualisation of London house prices from 1995 to 2011, with red indicating high and blue indicating low.	102
Figure 4.21 The structural attributes statistics.	103
Figure 4.22 The spatial distribution of housing attributes mapped using GIS.....	104
Figure 4.23 The spatial network model of London.	105
Figure 4.24 Greater London Public Transport Innovation projects between 1995 and 2011.	106
Figure 4.25 The Greater London closeness centrality with a radius of 20 kilometres.	106
Figure 4.26 Visualisation of the retail amenity using a 400-by-400 metre grid, where red denotes a higher density of retail and green denotes a lower density of retail.	107
Figure 4.27 Visualisation of Greater London secondary school average A-level scores, where red denotes higher scores and green lower scores (DOE, 2015a).	109
Figure 4.28 Greater London registered parks and gardens (English Heritage, 2014).....	110
Figure 4.29 Visualisation of the London jobs spatial distribution from 1998 to 2011, where red indicates a more jobs and blue indicates fewer jobs.....	111
Figure 4.30 The UK Postcode Unit, LSOA, MSOA, Ward and GLA visualised for Thamesmead in London.	112
Figure 4.31 Greater London Spatial Street Network dataset.	112
Figure 4.32 Visualisation of St-LAs in the Greater London area.	113
Figure 4.33 Postcode Units in Greater London.....	114
Figure 4.34 The UK Electoral and Census Area Statistics Ward.....	114
Figure 4.35 Super Output Area of Greater London.	116
Figure 4.36 Different housing submarkets used for Research Strand Three.....	116
Figure 5.1 This chapter focuses on the accessibility effects on house prices.....	118
Figure 5.2 Diagram showing how geometric accessibility and geographic accessibility overlap with each other.	120
Figure 5.3 Geographic and geometric accessibility.	121
Figure 5.4 The area the Greater London case study.....	127
Figure 5.5 Average house prices in Greater London between 1995 and 2011.	128
Figure 5.6 House prices in London in 2010, 2005, 2000, and 1995, visualised from red (high) to blue (low) using a constant colour range.....	129
Figure 5.7 London house price per square metre in 1995 (left) and 2011 (right).	130
Figure 5.8 House prices per square metre comparison between the JLE, the CTRL, the DLR and the London OG.	130
Figure 5.9 Transport projects in London between 1995 and 2011.	131

Figure 5.10 Accessibility measures mapped by a colour spectrum, where red denotes high accessibility and blue denotes low accessibility.	132
Figure 5.11 Closeness centrality measures for East London in 1995 (left), in 2011 (middle) and for the difference between the two years (right).....	132
Figure 6.1 This chapter focuses on the St-LA effect on house prices.	142
Figure 6.2 The Canary Wharf boundary (blue) overlaid with the MSOA boundary (red).....	143
Figure 6.3 Traditional administrative local area (left) and St-LA (right).	144
Figure 6.4 The steps in constructing the St-LA subgraph.	147
Figure 6.5 North London subgraph network.	148
Figure 6.6 Multilevel regression framework.....	151
Figure 6.7 Study area boundary.....	154
Figure 6.8 Visualisation of London house prices in 2011, with red indicating high house prices and blue indicating low house prices.....	155
Figure 6.9 The London OS Meridian street network.	156
Figure 6.10 Visualisation of St-LA membership in the Greater London Area.	157
Figure 6.11 Isle of Dogs area as denoted by LSOA, MSOA and ward boundaries.	157
Figure 6.12 London spatial house price clustering effect.	160
Figure 6.13 The visualised London St-LA subgraph map.	162
Figure 6.14 Scatterplot between inter-cluster house price deviation and inter-cluster distances. ...	162
Figure 6.15 Intra-class correlation coefficient comparisons.....	166
Figure 6.16 Goodness of fit comparison (AIC) across different local area units.....	167
Figure 6.17 Two abstract subgraph connected via a bridge.	168
Figure 6.18 Simulation of a random walker on the abstract graph.	168
Figure 6.19 Average number of steps required to jump between clusters.....	170
Figure 7.1 This chapter focuses on the housing submarket level.....	172
Figure 7.2 Street-based housing submarkets that combine street network attributes, spatial attributes and structural attributes (left) compared to traditional housing submarkets formed by spatial-structural attributes (right).	174
Figure 7.3 The submarket construction process.....	175
Figure 7.4 Multilevel hedonic price model framework	178
Figure 7.5 The Greater London study area.	182
Figure 7.6 The five local area units in Thamesmead.....	183
Figure 7.7 Weighted mean square error test summary.	185
Figure 7.8 Greater London St-HSs.....	186
Figure 7.9 Comparison between the St-HS and the ICC.	191
Figure 7.10 Visualisation of the St-LA, ward, MSOA, LSOA and postcode housing submarkets. Colours denote different housing submarkets.	192
Figure 7.11 The AIC comparison across the different housing submarket models.	193
Figure 7.12 London house prices per square metre for 1995 (left) and 2011 (right).....	194
Figure 8.1 Intra-city Hedonic Price Model	197
Figure 8.2 Space Syntax Integration (Closeness Centrality) and Gravitational Potential. Accessibility measures mapped by a colour spectrum, where red denotes high accessibility and blue denotes low accessibility.	198
Figure 8.3 Green Lanes normalised (1995-2011) house price per sqm.....	199
Figure 8.4 London Street-based Local Area in identifying the Isle of Dog. Colours denote different St-LA.	200
Figure 8.5 Random walker simulation.	201
Figure 8.6 Greater London St-HSs. Colours denote different housing submarkets.	202

Figure 8.7 Visualisation of the St-LA, ward, MSOA, LSOA and postcode housing submarkets. Colours denote different housing submarkets.	203
Figure 8.8 London house prices per square metre for 1995 (left) and 2011 (right).....	203
Figure 8.9 UK spatial network closeness centrality on the left and house price on the right.	208
Figure 8.10 Birmingham spatial network closeness centrality on the left and house price on the right.	208
Figure 8.11 Scatterplots between house price per square metre and spatial network closeness centrality in London on the left and in Birmingham on the right.....	209
Figure 8.12 Local area cores (green) and fuzzy boundaries (red).....	210
Figure 8.13 House price spillover simulations. a. House price simulation in step one b. House price simulation in step twenty	211
Figure 8.14 Active frontage score predicted by using a deep learning classifier. Source: Author.	212

List of tables

Table 3.1 Statistical comparison between community detection algorithms.	77
Table 4.1 Hedonic price model dataset and variables specification.	97
Table 4.2 The Land Registry house price dataset descriptive statistics.	99
Table 4.3 The Nationwide Building Society house price descriptive statistics.	100
Table 4.4 House Price descriptive statistics in 2011.	101
Table 4.5 The structural attributes statistics.	103
Table 4.6 Street and TfL spatial network datasets.	105
Table 4.7 The retail address points dataset.	107
Table 4.8 Secondary school dataset (DOE, 2015a).	108
Table 4.9 School average scores in 2011 (DOE, 2015a).	109
Table 4.10 Job statistics between 1998 and 2011 at the LSOA level.	111
Table 4.11 Street-based local area statistics.	113
Table 4.12 Postcode Unit statistics for Greater London.	114
Table 4.13 The electoral ward statistics for population and employment.	115
Table 4.14 The MSOA population and employment statistics (ONS, 2015).	115
Table 4.15 The LSOA population and employment statistics (ONS, 2015).	115
Table 5.1 The parameter space for the three accessibility measures.	123
Table 5.2 Specification of the hedonic model variables.	124
Table 5.3 London street-tube network model specifications from 1995 to 2011.	126
Table 5.4 OLS pooled regression models.	126
Table 5.5 Fixed-effect regression models.	126
Table 5.6 Joint accessibility models.	127
Table 5.7 Descriptive statistics for the variables used in the empirical analysis.	133
Table 5.8 Pooled regression model for radius infinity.	135
Table 5.9 Pooled regression model for radius 60 minutes.	135
Table 5.10 Pooled regression model for the three transport lines in London.	136
Table 5.11 Fixed-effect regression results for radius infinity.	136
Table 5.12 Fixed-effect regression results for radius 60 minutes.	137
Table 5.13 Joint accessibility fixed-effect regression model results.	138
Table 5.14 Variation inflation factor (VIF) measured for the six accessibility measures. VIF>10 is highlighted in red.	139
Table 6.1 Candidate models.	151
Table 6.2 St-LAs detected in Greater London.	156
Table 6.3 Local area statistics.	157
Table 6.4 Descriptive statistics for house prices and attributes.	158
Table 6.5 ANOVA Statistics. The results suggest house prices are more similar within local areas.	159
Table 6.6 North London St-LA inter-cluster house prices.	160
Table 6.7 Inter-cluster regression analysis results.	163
Table 6.8 Multilevel hedonic regression results.	164
Table 6.9 Comparison of the AICs.	165
Table 7.1 The candidate models.	178
Table 7.2 Local area housing submarket models specifications.	182
Table 7.3 Local area statistics.	183
Table 7.4 Descriptive statistics for house prices and attributes.	184
Table 7.5 Chow test summary (F-tests).	185
Table 7.6 Greater London St-HS summary.	186
Table 7.7 Individual housing submarket regression summary.	187
Table 7.8 Multilevel hedonic price regression model results.	189
Table 7.9 Street-based housing submarket goodness of fit comparison.	190

Chapter 1

Introduction

1.1 Background

“...value depends on rent, and rent on location, and location on convenience, and convenience on nearness” (Hurd, 1903. Page 13)

Understanding the economic value of space is an important topic in urban research, as space structures the nearness and farness for people and information. Despite the rise of technology and the alleged death of distance (Cairncross 1997), the demand for face-to-face interaction and agglomeration in the age of information appears to be greater than ever before (Glaeser 2011; Bettencourt 2012). The competition for the most connected space in the most connected city is reflected in the ever-higher real estate prices for central areas around the world. This research takes the spatial configuration approach to describe cities as spatial networks (Hillier and Hansen 1984). The key reason for this methodological adoption is that cities are not isolated entities (Hildreth 2006) but, rather, are complex systems (Wilson 2000) whose success in attracting and sustaining people and jobs depends on the linkages that exist within and between them. In addition, a city's fundamental advantages are its spatial configuration and the public good this configuration produces (Webster 2010). By retrieving the relative economic value of space, resources can be allocated more efficiently, well-designed places can be maximised, infrastructure costs are recoverable and urban designers are then able to objectively weigh between alternative designs. To facilitate these goals, value-capture instruments such as land pricing, taxes, facility pricing, externality pricing and neighbourhood pricing can be implemented. Therefore, it is important to understand the economic value of spatial configuration and the implication it has on the performance of cities.

One approach to retrieve this economic effect and value of spatial configuration is through the hedonic price approach. Using a hedonic price approach, the cost of a heterogeneous good, such as housing, can be broken down into its utility-bearing parts (Rosen, 1974). Since its introduction, this approach has become an established real estate method for pricing environmental goods, constructing housing price indices and as evidence in the development of welfare policies (Palmquist, 1989). Therefore, through a case study of Greater London, this research proposes the use of spatial configuration methods and the hedonic price approach as empirical strategies to enhance the analysis of the housing market and to capture the economic effects of spatial configuration. Thus, this research asks the following question:

What is the effect of spatial network configuration on intra-city house prices?

This chapter is set out as follows. The first section introduces the general research question. The second section describes spatial configuration as a measurement of urban form; this section also discusses the hedonic price approach as the empirical strategy. The third section introduces the three

empirical research topics concerning the effects of accessibility, the local area and the housing submarket on house price. The final section describes the thesis outline and concludes with a concise summary of the chapter.

1.1.1 Urban Rent Theory

The theory concerning space and rent can be traced back to Von Thunen's study in 1826 on the location of market places. Von Thunen found that agricultural activities that were the most productive and had the highest transport costs occupied locations closest to the market; however, agricultural activities that were less productive with a lower transport costs were located further away (Von Thunen, 1826). In an isolated context, this process creates a system of concentric rings where different agricultural uses radiate from the central market place. In 1946, William Alonso extended the Von Thunen agricultural model into the urban monocentric model. The urban monocentric model explained the centralisation of commercial activities in which house price diminishes as one moves further away from the central business district (CBD). Alonso's model is made operational through an assumption where the bidder who capitalises the most from the land can pay the highest rent in consuming it (Marshall 1890; Alonso 1946). Modern economic agglomeration theories add to this model by suggesting that the centralisation of commercial activities is explained not only by the reduction of transport costs to the CBD but also by the reduction in transport and information costs to all other businesses where knowledge sharing, matching, and learning can take place (Puga 2010; Krugman 1996). Alternative theories from urban sociologists put greater emphasis on historical and social factors rather than economic factors in explaining the shapes of cities (Park and Burgess 1925; Hoyt 1939; Ullmans and Harris 1945). These theories emphasised the fact that different social groups tend to sort, cluster, and expand at a zonal level, which results in neighbourhood clusters (Kain and Quigley 1975).

A key limitation in urban economic models and urban ecology models is the abstract representation of space and the simplification of processes (O'Sullivan 2004). Based on location theories, recent empirical work on land use and transportation models (Wegener 2004) have moved away from an abstract representation of cities to modelling the processes of cities with discrete administrative zones and actors. However, information can still be lost when urban form and urban geometry are not considered. The reason for this is that individuals move, experience, and interact within the city through its street networks and buildings.

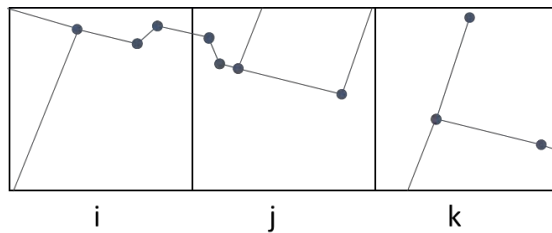


Figure 1.1 The limits of zonal representation of space.

Figure 1.1 shows a simple example that illustrates the importance of spatial representation. The example shows an abstract neighbourhood that is represented either by three contiguous zones **i**, **j**, **k** or by the street network graph that consists of a set of edges and a set of nodes. When using zonal contiguity, zone **i** is connected to zone **j** and zone **j** is connected to zone **k**. When using the street network, zone **i** and zone **j** are weakly connected (cognitively) and zone **j** and zone **k** are not connected at all. The advantage in spatial configuration research is it puts individual space and the relations between space at the heart of the method. A focus on individual space therefore makes spatial configuration an appropriate method to quantify the economic value of space.

1.1.2 Space Syntax Theory

Urban form, according to urban morphology traditions, studies settlements as stratified layers in cities across time. These layers include urban blocks, plots, buildings and streets (Conzen, 1964; Whitehand, 1990). One related school of thought, introduced by Bill Hillier and his colleagues at University College London (UCL), is space syntax, which focuses on studying the spatial configuration of urban layouts (Hillier and Hanson, 1984). This method argues that what happens in an individual space is fundamentally influenced by the relations between that space and the network of spaces to which it is connected. This spatial network perspective is used to address how individuals move, navigate and interact in cities (Hillier et al. 1993). Furthermore, space syntax as a theory postulates that spatial relations influence land use in cities via the activity generated by the grid (Hillier et al. 1997). This, in turn, changes cities' configuration through a feedback loop, where centrality expands, declines, shifts and diversifies under changing social and economic conditions (Hillier, 1999). At the heart of the space syntax theory is that humans preserve linearity or axiality when moving within cities (Conroy-Dalton 2001; Hillier and Iida 2005). An artefact of the axiality can be seen in the “deformed wheel pattern” of cities today ¹.

Methodologically, spatial configuration borrows from graph theory in representing spatial relations as graphs; nodes denote spaces and edges characterise the visible relations between spaces. Figure 1.2 illustrates the concept of using graphs to signify two different spatial configurations of the same house.

¹ Long axial line with commercial activities goes from the core to the edge connected by a rim where residential uses cluster in between (Hillier 2002). This outcome shares some similarities with Hurd's city expansion theory, which stresses the importance of a city's central area and axial (linear) growth (Hurd 1903; Hoyt 1939).

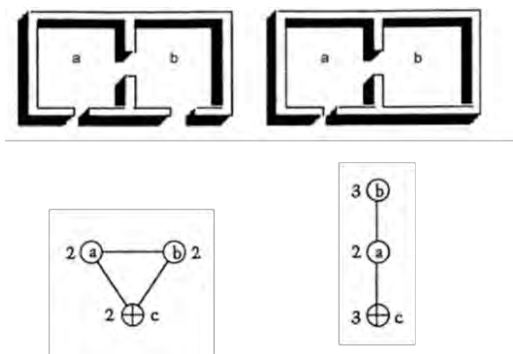


Figure 1.2 Two configurations of the same house conceptualised in a graph (Hillier and Vaughan, 2007).

From this representation of space, one can see how spatial relations afford different levels of integration and segregation. As shown in Figure 1, space 'b' may look the same in the two configurations, but it differs significantly as it sits in a different location within the overall network. When applied at the city level, the spatial configuration of the street network and its geometric relations construct similar spatial differences, where some areas are inherently less active (segregated) and other areas are inherently more active and central (integrated).

Focusing on the housing market, buyers are willing to pay various amounts for different types of space. Some buyers prefer to live in a more integrated space, while others would rather live in a more isolated space. This spatial scarcity and the differences in demand between integration and isolation create a market where spatial configuration structures the pattern of economic externality which can subsequently be priced (Webster 2015). Similarly, this can be extended to the neighbourhood or submarket level, where buyers are paying various premiums to live in different neighbourhoods that is also influenced by the configuration of the street network. This conjecture can be drawn from the theory of natural movement and movement economy, where space syntax theory provides a strong argument that spatial configuration is a lead cause for natural movement patterns, neighbourhood formation and land use pattern (Hillier 2005). As a result, this research intends to extend the space syntax theoretical argument by suggesting spatial configuration is producing spatial network accessibility effect and influencing the formation of community and housing submarket which in turn impacts upon residential location choice and structures house price pattern of a city. The following section describes the hedonic price model as an empirical approach to retrieving the economic effects of spatial configuration.

1.1.3 Hedonic Price Model as the Empirical Approach

One approach to reveal the effects and the economic value of spatial configuration is the hedonic price model (Ridker and Henning, 1967). The principle behind the hedonic price approach is that, holding all things constant, the intangible influence of spatial configuration can be discovered by observing real estate values. One can think of this concept by comparing two properties, each with nearly identical features, except that one property has one bedroom and the other has two bedrooms.

The price differential between the two is equal to the implicit price of the extra bedroom. Using this framework, the spatial network effects can be retrieved from the observed real estate values. This research specifically focuses on using the hedonic price approach on modelling intra-city house prices.

The hedonic price model is grounded in three theories. The first is the theory of spatial equilibrium (Glaeser, 2008; Von Thunen, 1826), which explains the trade-offs between transport cost and price. The second is the hedonic price theory, which explains the trade-offs between housing characteristics (as a bundle of structural, location and neighbourhood amenities) with price. The third is the theory of the housing submarket (MacLennan and Tu, 1994), where local submarkets have distinct embodiment in the supply and demand of a housing good. These three theories provide the foundation for applying an empirical framework that links intra-city house prices with spatial configuration features.

To summarise, there are several reasons for the adoption of the hedonic price approach. The first is that this method is a robust and well-established theoretical and methodological framework for estimating the effects and the economic value of spatial network configuration using observed house price. The second, is that this model only requires the transaction house price and the composition of the housing attributes in order to derive the spatial configuration effects on house price.

1.1.4 Greater London Case Study

The case study for the thesis is the Greater London housing market between 1995 and 2011. House prices in Greater London have risen dramatically and disproportionately over the past 15 years. This rise has been attributed to a combination of issues, including population growth, limited housing supply, policies limiting population dispersal, foreign speculative investments and a demographic transition in which the young middle class are flocking to central urban areas. This dramatic rise in house prices has subsequently led to a housing crisis, which not only affects the affordability (i.e. housing costs account for more than 15% of an entire lifetime of income) but also the productivity of the city (i.e. housing amounts to 25% of the national gross domestic product (GDP)). To alleviate this demand, 40,000-60,000 dwellings per year for the next 35 years must be constructed in London (GLA 2015). The key to solving these issues is either building more housing in under-supplied areas or increasing demand in under-performing areas via constructing more transport links. These solutions will benefit from further research on how the urban built form and its spatial configuration affect real estate value. This research is not only important for the Greater London area but also for the many growing mega-polycentric regions around the world, such as Sao Paulo, Brazil and Shanghai, China (Hall, 1996).

1.2 Empirical Strategy

This research proposes using a spatial configuration perspective to quantify urban form and the hedonic price approach as an overall empirical strategy for retrieving the economic value of the spatial form using Greater London as the case study. This research adopts the standard intra-city hedonic price model framework as the empirical strategy. Figure 1.3 illustrates the intra-city hedonic

price modelling framework, in which the price of a house can be deconstructed into its constituents, such as location attributes, structural attributes, neighbourhood attributes and the housing submarket in which it is situated (Freeman 1979). This research aims to contribute to this intra-city hedonic price model by adding geometric accessibility effects, street-based local area (St-LA) effects and street-based housing submarket (St-HS) effects into the model.

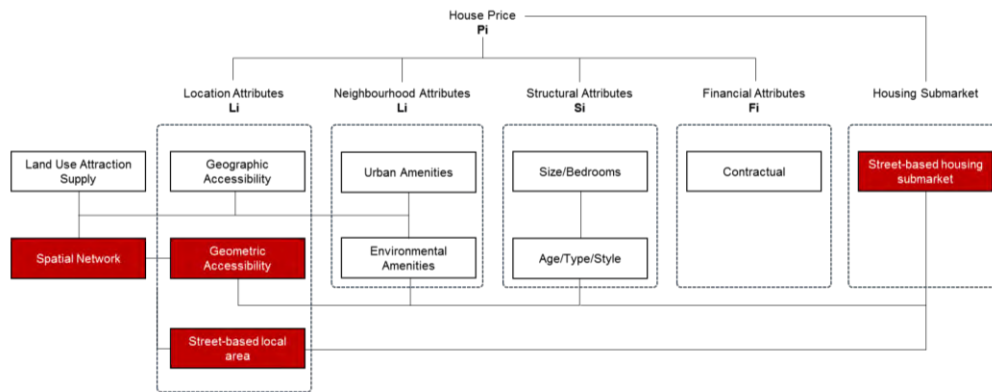


Figure 1.3 Intra-city hedonic price model. Spatial configuration effects highlighted in red.

1.2.1 Accessibility

The first strand of the argument focuses on the association between accessibility and house prices at the property level. Accessibility is defined as a measure of the potential interactions or relative proximity or nearness of individuals or places to all others (Hansen, 1949). Accessibility can generally be split into two broad classifications: geographic accessibility and geometric accessibility.

Geographic accessibility concerns the interaction between the functional attraction and the distances between attraction; geometric accessibility primarily concerns the spatial network configuration itself (Jiang et al, 1999). Existing results show strong associations between geographic accessibility and house price, using measures such as Euclidean distance to the central business district (CBD) in hedonic price modelling (Kain and Quigley, 1970). The logic is consistent with Alonso's monocentric city model (1964), where buyers are willing to pay more to live near central areas.

One limitation is that this approach assumes an endogenous definition of a CBD. This is an unrealistic assumption. For example, Heikkila et al. (1989) have found that house prices rise rather than fall with distance from the CBD in the Los Angeles Metropolitan Area. This conclusion led to the use of polycentric employment accessibility models which moved away from the idea of all economic activity being concentrated in a single, dimensionless point to a more heterogeneous distribution of employment. This idea is in-line with Ullman and Harris's (1945) multiple-nuclei model in geography and aligns with the use of polycentric accessibility measures from urban geography (Adair et al. 2000; Wilson 1970). A second limitation is that these geographic measures of accessibility have focused less on general accessibility effects, as explained by Webster (2010). This limitation has led to an interest in the field of space syntax, which has methods to allow the quantification and valuation of general accessibility. This argument is supported by empirical research, where significant positive

associations were found between spatial network accessibility and the Council Tax band in London (Chiaradia, Hillier, Barnes, Schwander, 2012a; 2012b), house prices in Cardiff (Xiao et al. 2015; Narvaez et al. 2014; Narvaez 2015), house prices in Shanghai (Yao and Karimi 2015) and house prices in London (Law et al. 2013).

Several research gaps remain. The first concerns the lack of research in generalising the results. Focusing on a single case study reduces the generalisability of the results and the extent to which different spatial contexts might value accessibility differently (Law et al. 2017a). The second concerns the lack of research on general accessibility effects (geometric accessibility) and the lack of comparison between geographic and geometric accessibility measures; this gap is the focus of the first research strand.

Research strand one question: To what extent are measures of geometric accessibility associated with intra-city house price variations? how does geometric accessibility compare with geographic accessibility measures when associating with house price?

1.2.2 Local Area

An important but under-explored topic within the field of housing studies is the definition of local area. The concept of locality or neighbourhood is complex which encompasses spatial, historical, socio-economic and perceptual constructs that change and overlap according to the geographical scale and point in time (Galster, 2001). Stemming from Kearns and Parkinson's (2001) definition, this research defines the local area as a geographical unit that is larger than a home area but smaller than a district. Much existing research focuses on using census tract administrative output area as units to estimate effects on house price. One criticism of this census tract administrative output area, however, is its 'arbitrary' and 'ad hoc' definition (Orford, 1999; Jenks and Dempsey 2007). This arbitrariness can be mainly attributed to the statistical method employed, the anonymity of the data and the historic dependence, which can cause inconsistent results (Goodman, 1979; 1985). To counter these issues, the proposed study uses an Street-based-local-area (St-LA) approach, which can more accurately capture subtle differences in the urban environment. As a result, this strand of the research compares these two local area specifications to answer the following research question:

Research strand two question: To what extent do St-LAs, as defined by the topology of the street network, associate with house prices? Secondly, how do St-LAs compare with administrative local area units when correlating with house prices?

1.2.3 Housing Submarkets

Rosen (1974) has described an underlying housing market model in which the market settles on a set of prices when supply meets demand. A criticism of the hedonic price model is that market-clearing prices are not expected to equalise when there are various implicit prices across local markets (MacLennan and Tu, 1996; Leishman, 2009). Different local markets will demand a variety of bundles,

which inevitably leads to varying prices for different submarkets (Goodman and Thibodeau, 1998). This segmentation unavoidably ushers in the identification of housing submarkets, which is one of the most discussed topics in housing literature.

Grisby et al. (1987) defined housing submarkets as units that are reasonable substitutes for one another but relatively poor substitutes for units in other submarkets. This homogeneity can be determined *a priori*, through e.g. real estate agents, or it can be empirically driven, for example, by either spatial or structural factors, or a combination of both. Despite a general consensus on the existence of the submarkets (Watkins, 2001) and the statistical tests used to infer them through differences in implicit prices (Schnare and Struyk, 1976), there are general disagreements concerning their identification (Watkins, 2001; Bourassa et al., 1999). For example, if submarkets can either be determined *a priori* or empirically driven, should they be spatial or structural, or should a housing submarket be well defined or fuzzy, with overlapping boundaries (Helbich et al., 2013)?

One subtopic that has been rarely discussed concerns the geography of submarket identification. To define housing submarkets, the majority of existing researches use census tract output areas as building blocks. The general procedure is to aggregate administrative units with similar implicit prices into individual submarkets through a clustering procedure. However, the use of arbitrary administrative units can lead to unclear submarket definition and similarly inconsistent results. This thesis consequently propose that Street-based housing submarkets (St-HS) as defined by St-LA can described spatial housing submarket more accurately than traditional housing submarket using administrative local area. The reason for this is that housing submarkets are also driven by spatial configuration attributes. This research strand thus identifies the following research question:

Research strand three question: To what extent are St-HSs comparable to traditional census tract-based housing submarkets when correlated with house prices?

1.3 Thesis Outline

Chapter 1: Introduction

Chapter 1 addresses the general background and motivation of the research, the research questions and the structure of the thesis. In particular, it addresses the following: What is my research question? How will it be answered?

Chapter 2: Review of the Hedonic Price Model

Chapter 2 provides a review of the hedonic price model literature as an approach to identifying the effects and implicit prices of spatial configuration from observed house prices. The first part introduces spatial equilibrium theory, hedonic price theory and submarket theory as the foundations for modelling intra-city house prices. The second section describes the use of the four key empirical topics in examining intra-city house price variations, namely structural characteristics, location

characteristics, neighbourhood characteristics and housing submarkets. The last section identifies the three corresponding empirical topics to be studied in the analytical chapters.

Chapter 3: Spatial Network Methods

Chapter 3 introduces spatial network methods to respond to the three research questions set out in the previous chapter. This chapter includes a description of spatial network centrality for measuring geometric accessibility and a description of community detection techniques that can be applied to spatial dual graphs in order to identify St-LAs. Furthermore, this chapter incorporates a description of standard k-means clustering algorithm to identify St-HSs.

Chapter 4: Research Framework, Case Study and Dataset

Chapter 4 describes the use of the hedonic price regression model to answer the research question, which involves testing the association between street-based geometric properties and house price across three scales: the property level (Chapter 5), the local area level (Chapter 6) and the submarket level (Chapter 7). This chapter also introduces London as the case study and the various datasets that are used for the three analytical chapters.

Chapter 5: Geometric Accessibility

Chapter 5 is the first analytical chapter. This chapter studies the extent to which the geometric accessibility effect is significant when associated with house prices. The first section describes geographic and geometric accessibility measures. The second part introduces the fixed effect regression model, which estimates the effects of accessibility on house prices. The third section reports the results of the regression model and discusses these findings.

Chapter 6: Street-based Local Area

Chapter 6 is the second analytical chapter. This chapter explores the extent to which the St-LA effect is significant when associated with house prices. The first part describes the local areas that will be compared, which includes the St-LA and three other administrative census tract output areas. The second section introduces the multilevel hedonic price regression model, which estimates the local area effects on house prices. The third part reports the regression model results and discusses the findings.

Chapter 7: Street-based Housing Submarket

Chapter 7 is the third analytical chapter. This chapter investigates the extent to which St-LAs improve housing submarket identification when associated with house prices. The first part describes the standard, K-means clustering method used to identify the housing submarkets. This includes the St-HS and the census tract output area housing submarket. The second section reintroduces the multilevel hedonic price regression model, which estimates the effects of the housing submarkets on house prices. The third part reports the regression model results and discusses the findings.

Chapter 8: Conclusion and Ways Forward

Chapter 8 summarises the empirical results from the three analytical chapters. This chapter also discusses the key findings, research limitations and suggestions for future work in this area, which includes additional tests across geographical regions, the identification of fuzzy local areas, the examination of the house price spillover effect across time and a pilot study applying deep learning methods in extracting the economic value of urban design.

1.3.1 Summary

To summarise, this research proposes the use of spatial configuration methods and the hedonic price approach to better understand the economic value of spatial configuration and to enhance the analysis of the intra-city housing market in Greater London. This research aims to study spatial configuration effects on house prices across three levels: the accessibility level, the local area level and the housing submarket level. By understanding the spatial configuration effects on house prices, clear mechanisms and policies can then be proposed to encourage better design and to improve resource allocation. On one hand, this research borrows methods and concepts from regional science, real estate economics and space syntax to grasp the economy of cities. On the other hand, this research hopes to better understand the intangible value of space and design. The next chapter provides a more in-depth introduction to the hedonic price approach as a theory and technique for revealing both the economic value and the effects of spatial configuration.

Chapter 2

Review of the Hedonic Price Model

2.1 Background

This chapter reviews the hedonic price model as an approach to retrieve the effects of spatial network configuration from the observed house prices. The chapter is organised as follows. The first part describes the spatial equilibrium theory, the hedonic price theory and the housing submarket theory as the foundation for the hedonic price model. The second section describes the empirical approach and includes a discussion on the variable set, the functional form, and the regression specifications. The third part summarises the assessment and, from the spatial configuration perspective, identifies the research gaps in hedonic price models. This review is mapped in Figure 2.1.

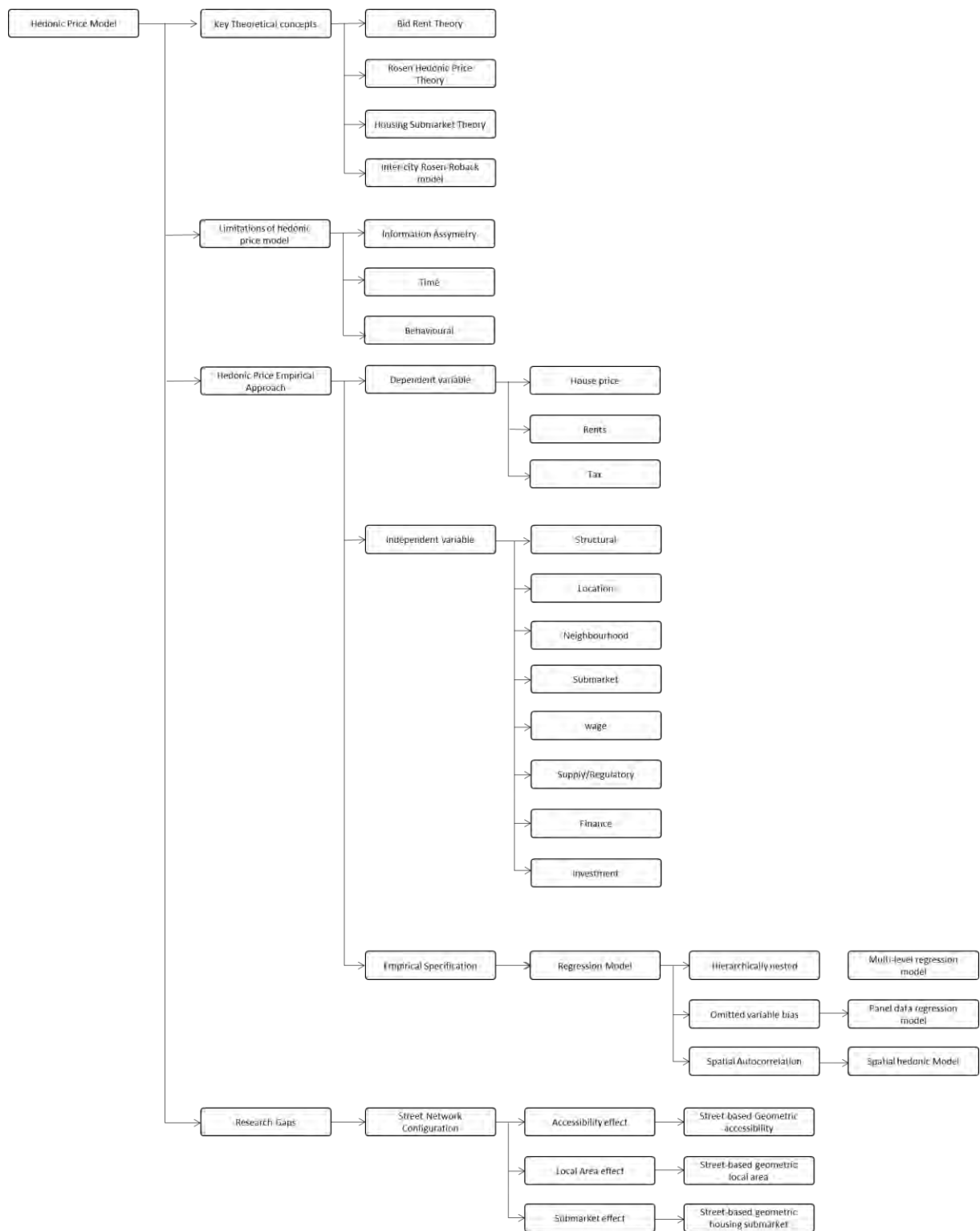


Figure 2.1 Map of the review of the hedonic price model.

2.2 Key Theoretical Concepts

*‘While I believe that no one can make sense of cities without the tools of economics, I also believe that no economist can make sense of cities without borrowing heavily from other disciplines’
(Glaeser 2008, p.32).*

Following Rosen’s (1974) economic framework and Ridker and Henning’s (1967) seminal empirical study, the hedonic price method has become one of the most popular approaches to the valuation of intangible goods, such as school quality, noise or pollution (Black 1999; Ridker and Henning, 1967), constructing quality of life indices (Rosen, 1979; Roback, 1982) and as inputs in land use and transportation modelling (Waddell, 2002; Lochl and Axhausen, 2010). The next section describes the key theories in the hedonic price approach, as outlined in Figure 2.2.

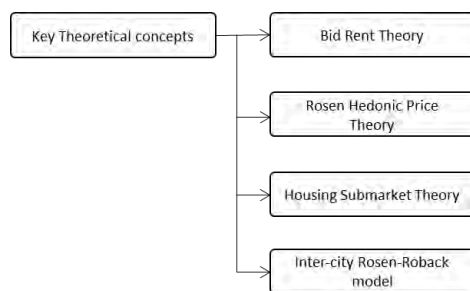


Figure 2.2 Key theoretical concepts for the hedonic price model.

First, borrowing from urban economics, the concept of spatial equilibrium and the bid rent monocentric model are discussed. Second, hedonic price theory is described as a theoretical and empirical framework for deconstructing the housing bundle as a set of locational, structural and neighbourhood attributes. Third, the housing submarket theory is described. Last, the research provides a brief discussion on the inter-city hedonic price model, which is not used in this thesis.

2.2.1 Spatial Equilibrium and the Monocentric Bid Rent Model

This study begins by describing the concept of spatial equilibrium in cities (Glaeser, 2008). Spatial equilibrium, or the spatial trade-off theory, originated from Von Thunen’s (1826) study on the location of marketplaces. He found that agricultural activities that were both most sensitive to transport costs and utilised the least amount of land were located near the market, whilst agricultural activities that were less sensitive to transport costs and required more land were located furthest away. Following the work of Von Thunen (1826), Alonso (1964) developed the monocentric model to explain the centralisation of business and commercial activities, where density, land rent and house prices diminished with increased distance from the city centre. This model operates via a bidding process, where the person with the highest bid acquires the right to the land. The typical bid rent diagram is illustrated in Figure 2.3. The y-axis is the rent per square metre, and the x-axis is the distance from a predetermined central place. The three overlapping, downward sloping bid rent curves represent the rents that a bidder from each respective land use (retail, manufacturing and residential) is willing to pay as a function of the

distance away from the centre. The outer envelope of the three curves (orange) represents the highest bids and, therefore, the land use for each location. This model creates the famous concentric land use pattern, where a centrality-rich land use offering the highest rents, such as retail, is situated in the CBD, and a less centrality-rich land use offering lower rents, such as residential, is located furthest away.

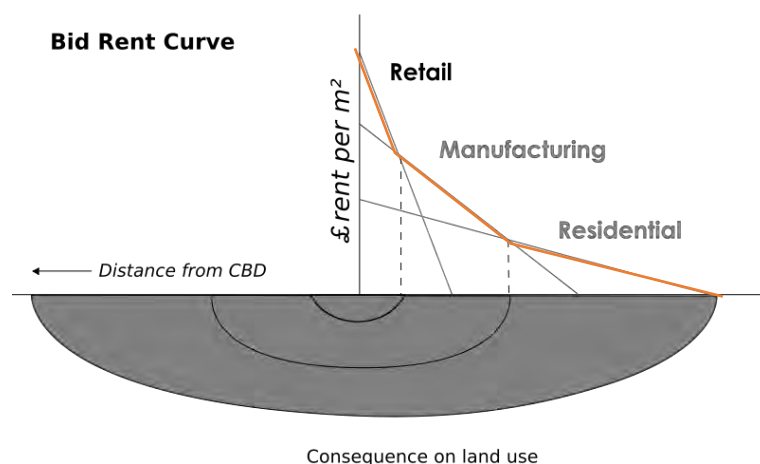


Figure 2.3 The bid rent curve. Credit: by SuzanneKn at Wikipedia (Wikimedia Commons).

From a residential location perspective, this model generates an inverse relationship between house prices and the distance to central places. In Alonso's monocentric model, the transportation network is assumed to be unimodal and isotropic. Dwelling and household preferences are also homogeneous. The key to the theory of spatial equilibrium is that income, amenities, housing costs and transport costs are equalised. $Income + Amenities - Housing Costs - Transportation Costs$

The essence of the monocentric model is its simplicity and elegance in both explaining land rent through transport cost and showing how simple processes can explain much of the shape of cities today. This theory is central to explaining accessibility effects in hedonic price models and is an essential theory in land use transportation models (Batty, 1976; Lowry, 1964; Wegenaar, 1994).

The monocentric model has some well-known limitations. These limitations include the assumption of a predetermined CBD location and the supposition that transport costs are the sole factors in explaining house prices. Both of these assumptions are only realistic in abstract economic models. The first assumption has relaxed through the development of polycentric models of the city (Glaeser and Khan, 2001; Henderson and Mitra, 1994). The second assumption requires the incorporation of other factors into the house price model which leads to the discussion of the hedonic price theory below.

2.2.2 Hedonic Price Theory and Intra-city House Price Model

Court (1939) and Griliches (1967 and 1971) were the first authors to coin the term 'hedonic price models' and to use regression techniques to deconstruct a consumer good. In a more formal sense, Lancaster

(1996) laid the foundations for the hedonic price approach in 'A New Approach to Consumer Theory' by stating that the demand of a good is based on the utility that is linked to a good's characteristics. The neoclassical concept of utility maximisation is a key assumption in the hedonic price theory, whereby individuals are assumed to make decisions that maximise their utility. Rosen (1974) advanced this concept further to develop an urban economic theory called hedonic price and implicit markets, stating that a differentiated good, such as housing, is made up of utility-bearing characteristics that cannot be separated. If a household is in equilibrium, the marginal implicit price for an attribute will be equal to the marginal willingness to pay for the attribute. (Freeman 1979). Rosen's insight was that, given consumer preferences and income, the benefit of improving any one part of the bundle must be offset by the costs of the additional expenditure. This model allows intra-city house price to be deconstructed into a bundle of housing characteristics, such as structural characteristics, neighbourhood amenities and location accessibility (Freeman 1979). Empirical detail is spelled out in section 2.3.

2.2.3 The Housing Submarket Theory

A criticism of the hedonic price model is that market clearing prices are not expected to equalise. Buyers from different local markets demand different bundles and, therefore, are willing to pay different prices (Goodman and Thibodeau, 1998). For example, families with children value the school catchment area differently than those without children. This is known as the housing submarket hypothesis, where various submarkets have different implicit prices (Maclennan and Tu, 1996; Leishman, 2009). The housing submarket theory forms the third mechanism, whereby a city's housing market is composed of different local housing submarkets. Section 3.6 provides a more detailed discussion on the housing submarket as part of hedonic price modelling.

2.2.4 The Rosen-Roback Inter-regional Model

'Workers require a compensating wage differentials to live in a big, polluted or otherwise unpleasant city, the firms in that city must have some productivity advantage to be able to pay the higher wages.'

Jennifer Roback 1982

The Rosen-Roback model takes the original intra-city hedonic price model further by examining inter-city house price variations. Following the hedonic price theory, Rosen (1979) presented a hedonic wage model, where the quality of life ranking was constructed as a bundle of wages, rents and amenities. This model led to Roback's (1982) research, which formed the basis of the Rosen-Roback inter-regional model, where metropolitan residential location choices were based on the combined effects of wages, rents and amenities, whilst holding commuting costs constant. A household's decision to live in any given city is based on income plus the amenities provided. Various extensions of this classical model have been proposed, such as relaxing the assumption on the location of firms and transport costs (Berger, Blomquist and Waldner 1987). The inclusion of income, the labour market and migration makes the model more complex than the intra-city hedonic price model (Glaeser, 2008). As this research primarily deals with intra-city house price variations, a full description of this type of model is not within the scope of this review.

2.2.5 The Limitations of the Hedonic Price Model

The hedonic price model is logical and consistent, but it is far from perfect (Freeman, 1979). Its limitations can be separated into neoclassical economic assumptions and behavioural economic factors. To interpret the hedonic price model as a measure of demand (marginal willingness to pay) for the attribute requires each household to satisfy the equilibrium condition. In a neoclassical economic model, for equilibrium to be fully achieved, all buyers and sellers are utility-maximising agents with perfect information and zero transaction and search costs, where the equilibrium price adjusts instantaneously (Freeman 1979). This abstract model is clearly not a perfect representation of the real estate market, as the housing market takes many years to clear (DiPasquale and Wheaton, 1994). First, the high transaction costs relative to the savings suggest that a quick return to the equilibrium price via short selling is unrealistic. Second, the high search costs also create the conditions for imperfect information. This lack of information is certainly changing with the rise of open data and web services such as Zoopla. However, as a property is inherently unique both in its location and structure, this asymmetry of information continues to exist, as a buyer or seller is not able to visit every single property.

In the field of behavioural economics, individuals are often influenced by others, and decision-making can be irrational with limited information (Kahnman, 1979; Shiller, 2000). Leamer (2007) has suggested that changes in house prices could stem from optimism in a buyer's market and pessimism in a seller's market. Shiller (2000) has further argued that house price differentials can be explained by irrational expectations and self-fulfilling narratives rather than by fundamental economic factors. Residential location choices are also often influenced by the buyer's social network (Tuononen and Law, 2017). These behavioural factors clearly influence house prices.

To be fair, these criticisms are not exclusive to housing price research, but rather to the field of economics. As Freeman (1979) has noted, the question isn't whether the model is perfect but rather if it provides useful knowledge. In another word, limitations in the model do not render the technique invalid for empirical purposes. Though, it is important to cautiously study these models, where associations cannot be thought of as causation. However, when associations are repeatedly found with different datasets and settings, the hypothesis that property values are associated with the joint housing characteristics can generally be supported (Muth and Goodman, 1998). The hedonic price approach offers many advantages, as it only requires the house price, which is the composition of the housing attributes, to derive the marginal attribute prices. To conclude this subsection, the hedonic price model offers an approach that allows housing to be deconstructed into a bundle of housing characteristics. The empirical strategy of the model and its various topics are discussed in the next section.

2.3 The Empirical Approach of the Hedonic Price Model

'In buying housing, families jointly purchase a wide variety of services at a particular location. These include number of square feet of living space, different kinds of rooms, a particular structure type, an address, accessibility to employment, a neighbourhood environment, a set of neighbours, and a diverse collection of public and quasi-public services including schools, garbage collection, and police protection.'

Kain 1970

Buyers' preferences in the housing market are often examined through two empirical approaches: the revealed preferences approach, which utilises empirical data, or the stated preferences approach, which uses structured interviews. Hedonic price modelling uses revealed preference methods to estimate the implicit prices of housing characteristics from the observed sold house prices (Rosen, 1974; Sheppherd, 1999). Discrete choice modelling utilises stated preference questionnaires to determine consumer's residential location choices (McFadden, 1977). This research focuses on the former due to the availability of house price data in the UK.

Rosen (1974) has proposed a two-stage empirical approach for the hedonic price method. In the first stage, house price **P** is a function of the independent variable **X** = {**x**₁, **x**₂, ..., **x**_n};

$$HP = f(X)$$

P is the price of a property

x represents the independent variable

Equation 2.1

The implicit price for the attribute can be determined by the first-order condition, where:

$$\frac{\delta X}{\delta P}$$

Equation 2.2

One can think of this concept as a comparison between two properties, holding all other features of the dwellings constant, where one property has one bedroom and the other has two bedrooms. The price differential between the two is equal to the implicit price of the bedroom. This method allows the house price to be deconstructed into its constituent parts and valued separately.

Rosen (1974) proposed that, in the second stage, a consumer's demand function for each attribute can be estimated, where the quantity supplied for the attribute **Q** is a function of its implicit price **P** and of socio-economic attributes such as income **Y** and age of the buyer, etc. (Freeman, 1979).

$$Q_d = F[P(Q), Y, age, \dots]$$

Equation 2.3

The second stage in estimating the demand function is difficult to carry out, as it requires full

identification of the buyers' and sellers' socio-demographics and preference data. Rather than approximating the demand function of an attribute, most research consequently focuses on the first stage of the hedonic price model in estimating the marginal implicit prices based on the changes in housing attributes (Goodman 1979). Estimating the implicit price does not provide a complete picture; instead, it provides useful evidence for the relative value of an intangible good. The first stage of the hedonic price model is also the focus of this research. Figure 2.4 illustrates the hedonic price regression model, which is split into the dependent variable set, the independent variable set and the empirical specification.

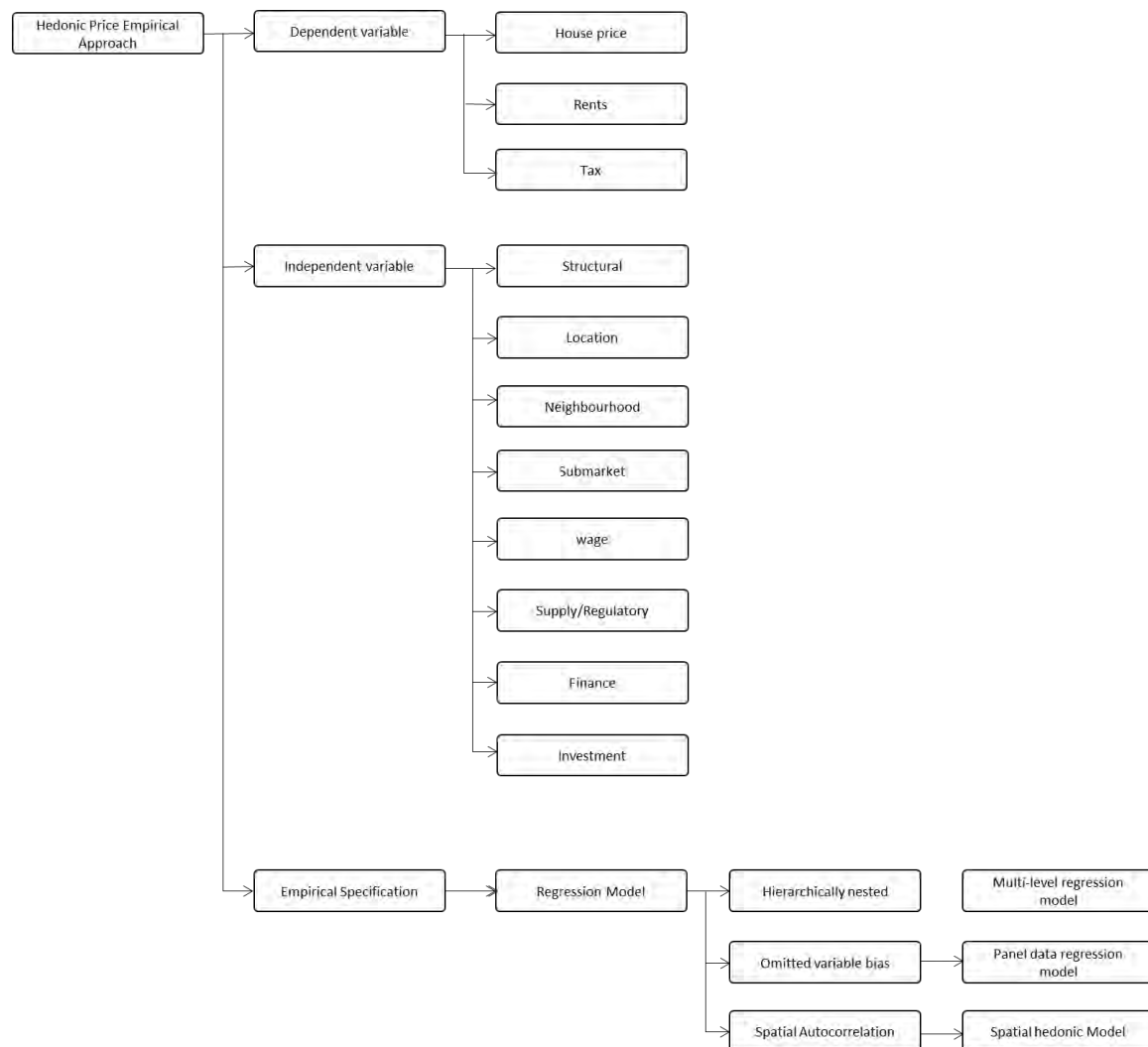


Figure 2.4 The hedonic price empirical approach.

2.3.1 The Dependent Variable: House Prices

The dependent variable in a hedonic price regression model can either be the sold price of the property, the rent of the property or the amount of taxes that can be incurred from the property. This research has selected the sold house price dataset as the dependent variable due to the availability of data and as the key variable used in previous research. There are, in general, two types of geographies for house prices. The first geography uses the census tract local area (Ridker and Henning 1967). The use of

census tracts suffers from interpretation problems due to heterogeneity within the geography. For example, the same census tract may contain both a large four-bedroom house and a flat. However, these types of data also contain a large array of socio-economic variables. The second type of geography is the more detailed postcode and property-level house price dataset (Straszheim, 1975; Goodman, 1978; Schnare and Struyk, 1976). This is an unbalanced dataset, where not every postcode has a transaction. Due to the large samples, concerns about sampling bias are subsequently reduced. As a result, this study selects the latter geography for this research.

2.3.2 The Independent Variable Set

Numerous independent variables are often included in empirical hedonic price studies. This research primarily focuses on variables that influence intra-city house prices, such as structural, locational, neighbourhood variables and the submarket the property is situated in (Freeman 1979; Watkins 2001). Variables like wage and housing supply are important factors pertaining to inter-city hedonic price models. Inter-city hedonic price model variables are briefly discussed at the end of the empirical studies review.

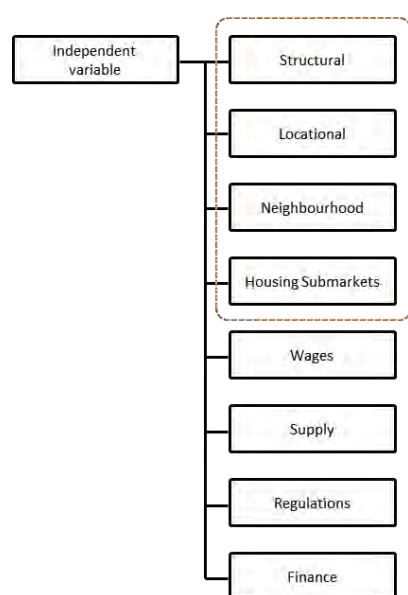


Figure 2.5 Intra-city hedonic price independent variables.

Several empirical meta-studies on hedonic price models have been published under different topics (Smith and Huang, 1995; Malpezzi, 2003; Sirman et al., 2006), including general reviews of hedonic price model variables (Sirman et al., 2006) and specific reviews focussing on environmental variables (Walter and Schlapfer 2010) and housing submarket variables (Watkins 2001). For example, Sirman et al. (2006), from 120 studies, have found that four categories of variables are dominant in hedonic price studies, including structural, locational and neighbourhood variables. The most dominant classifications are structural variables such as age, square metre, the number of fireplaces; locational variables, such as the distance to the CBD; and neighbourhood variables, such as school quality. The study also found that key variables such as accessibility, size, age, the number of bedrooms and the school district, are

significant more than 80% of the time. Appendix A contains charts produced by the author using this meta-study. To summarise, house prices are generally considered as functions of structural, locational, neighbourhood and submarket factors. A more detailed discussion concerning these four topics is provided in the next section. This paper also discusses the extent to which spatial configuration approaches are considered for each of these topics.

2.3.3 Structural variable

The influence of structural variables on house prices is well documented. The reasons are clear; for example, buyers pay a higher price for a larger quantity of a structural feature, such as floor area or fireplaces. These variables are commonly separated into internal features, such as size and the number of bedrooms, bathrooms, fireplaces, garages, and external features, such as lot size, the quality of the building, and the age of the building (Sirman et al., 2005; Kain and Quigley, 1970). Kain and Quigley (1970) have found that internal features have as much of an effect on house prices as external features. They also showed that, in the US, the age of the structure correlates negatively with house prices. In contrast, this effect is reversed where older buildings are more appreciated in European cities. This difference shows that building such models require careful examination and local domain knowledge for interpretation. Empirically, a property's structural attributes such as its age and its location are inherently related; there are more older buildings in the centre than in the suburbs. This is evident by the high degree of multi-collinearity between structural and location features and should be carefully considered in the estimation of the hedonic price model.

2.3.3.1 Structural variable limitation

One dimension that is less documented in the literature on structural variables is the extent to which the aesthetic architectural quality both internally and externally, rather than the functional quality as mentioned in the previous section, can influence house prices. This limitation can be attributed to the difficulties in identifying robust measurement instruments. Previous studies concerning the aesthetic dimensions of architecture have either focussed on using heuristics as proxy for architectural quality, such as the age of a building; or asking a panel of experts for a rating of a building's architectural aesthetic (Vandell and Lane 1989; Hough and Kratz 1983). Therefore, there are research gaps concerning these approaches.

Second, there is also a lack of research regarding how these design features produce utility for the users. For example, the floor-to-ceiling height of a home influences both the volume of a home and the amount of natural light that filters into each room. This design feature, therefore, produces functional, aesthetic and health utilities. Third, the architectural quality of a space is often a complex combination of multiple architectural features. An overlooked feature that encompasses this complexity is the configuration of a home regarding both the proportions and relations between rooms. For instance, what is the proportion or size of a particular room, such as the kitchen, and what is the relation between the room to all other rooms? How is the configuration for instance within a home valued? These topics are beyond the scope of this study; this research solely focusses on

spatial configuration within the public space rather than in the private realm. Thus, structural variables are considered controlled variables in the analytical chapters and are not extensively discussed in this thesis. However, there are opportunities for future research to study the extent to which architectural quality and spatial configuration of the private realm can influence house prices using the hedonic price approach.

2.3.4 Location Variables

In the hedonic price approach, estimating the marginal willingness to pay for location differentials or accessibility is an important topic. The variable is based on the monocentric model, where buyers trade-off between transport and rent. The variable is traditionally estimated in the form of Euclidean distance to the CBD (Alonso, 1964). The following section discusses how access to employment, public transport and private transport are used in hedonic price studies. This research distinguishes location variables as accessibility effects at the city-wide level, rather than accessibility effects at the neighbourhood scale which will be discussed in the next sub-section.

2.3.4.1 Access to Employment

Euclidean distance to the CBD is by far the most common city-wide accessibility measure in hedonic price modelling (Kain and Quigley, 1970; Osland et al., 2007; McMillen, 2004). This spatial separation accessibility measure assumes a homogeneous house price gradient that uniformly declines from a centre point of employment. Variations include imposing an asymmetric price gradient that varies according to slope and direction (Coulson, 1991). A limitation to this approach is that it assumes a monocentric structure of cities, like those of London or Chicago. However, this is not always the case. Heikkila et al. (1989) have found that house prices rise rather than fall with distance from the CBD in the Los Angeles metropolitan area, which contains multiple centres. More recent models acknowledge the polycentric structure of cities and use measures such as distance to multiple employment centres (McDonald and McMillen, 1990) and gravity-type accessibility measures (Ahfeldt, 2011) and singly constraint spatial interaction-type accessibility measures (Adair et al., 2000). These methods move away from the idea that all economic activity is concentrated in a single point to a more heterogeneous distribution of employment. Recent research has also shown this type of accessibility effects can differ across housing submarkets (Adair et al., 2000) and across geography (Law et al. 2017a).

2.3.4.2 Access to Public Transport

The next topic is related to public transport accessibility (Baum-Snow & Kahn, 2000; Cervero and Duncan, 2001; Gatzlaff & Smith, 1993). Debrezion, Pels and Rietveld (2007) have provided an extensive meta-analysis on this strand of research. Early research from Dewees (1978) and Bajic (1983) has found that new subway lines have had a positive effect on house prices in Toronto. Similar transport premiums on house prices were also found for the Philadelphia SEPTA system (Voith, 1993), the Boston MBTA system, and the London Underground system (Gibbons and Machin 2005). This effect can be observed for different types of public transport technology. For example, recent research from

Cervero (2011) and Munoz and Raskin (2010) has found that the Bus Rapid Transit (BRT) systems also exhibited a similarly positive impact on house prices.

The influence of public transport on house prices is similarly complex. Such complexity can be seen in the study by Lewis-Workman and Brod (1997), which compared the subway systems in San Francisco and New York City. These authors have identified a decline of approximately \$1,578 for every 100 feet that an individual moved away from a San Francisco BART station and a decline of about \$2,300 for every 100 feet one moved away from a New York MTA station. Holding income constant, this conclusion suggests that the effects of public transport on house prices clearly differ across geography. Henneberry (1997) found complex effects from the Sheffield Supertram; higher house prices were observed near the tram before construction, but no effects were observed while the system was in operation. These results show that the capitalisation of public transport is a complex process and can enter the market's life cycle at different times.

2.3.4.3 Access to Private Transport

Similar complex effects have been found with private transport projects, such as the motorways. For example, Boarnet and Chalermpong (2001) have found that new motorway projects have a positive effect on house prices in California. However, the maximum house price appreciation seems to occur at a moderate distance from the motorways (Chernobai, Reibel and Carney 2011). These results suggest that there is a trade-off between being too close to a motorway (noise-externality) and being too far away (accessibility-externality). Accessibility effects clearly differ between time and geography and are also related to travel behaviour and transport technology.

2.3.4.4 Location variable limitation

There are several limitations to the use of accessibility measures as location variable in hedonic price models. One topic is that the influence of accessibility can differ both geographically and across time. For example, McMillen (2003) has found strong evidence of this impact in the Chicago CBD, where the effects of employment accessibility on house prices were significant before 1980, insignificant in 1980 and significant again in 1990. Recent research from Law et al. (2017a) has found related evidence where the effect of accessibility varies geographical for UK cities. Similarly, a study by Adair et al. (2000) has found evidence that employment accessibility is more significant for the lower income submarkets than for other socio-economic groups. This fluctuation can be explained by the changes of the social geography of the residential location but also by the fact that different demographic groups in different economic structure is likely to exhibit different value towards accessibility (Law et al. 2017a). An important finding of these studies is that on average, accessibility variables are significant. However, rather than being generalisable, these effects are geographically dependent and are bound by an area's socio-economic, demographic, mobility and historical factors at that point in time.

Second, despite the large body of research concerning accessibility effects on house prices, most studies have focussed on using simple spatial separation (e.g. distance to the CBD) and spatial

opportunities (e.g. the number of shops within 800m) as measures of accessibility (Des Rosier et al., 1996; Handy and Niemeier 1997). Only more recently have gravity models been used (Adair 2000; Ahfeldt 2005). Geometric accessibility measures such as spatial network centrality and connectivity have only been recently considered (Xiao et al., 2015; Law et al., 2013; Nase et al, 2013). There is, consequently, a research gap concerning the comparison between different accessibility specifications in hedonic price models. There is also little research in combining both commuting behavioural data and accessibility measures into a composite measure for modelling location variables on house price (Theriault et al. 2005). This topic is further discussed in the first empirical chapter.

2.3.5 The Neighbourhood Variable

Bartik and Smith (1987) have defined amenity as the positive and negative contributions of location-specific goods and services at a neighbourhood scale. More simply stated, when an individual is purchasing a home, they are not only buying its structural features or its access to work, but also the amenities that the neighbourhood provides. Some amenities, such as air quality, are measurable whilst others, such as the beauty of a neighbourhood, are intangible. Three topics related to the neighbourhood amenities variable include environmental amenity, urban amenity and neighbourhood effects. Environmental amenity refers to the utility or benefits derived from nature, such as oceans, mountains and parks. Benefits are provided in terms of recreation, health improvement and psychological well-being. Urban amenity refers to the utility or benefits derived from accessing schools, shops and safety. Benefits are accumulated from economic attainment, physical safety, education and recreation. These types of amenity effects are called first-order effects, as the amenities directly affect the homeowner. Lastly, neighbourhood effects refer to the second-order indirect effects accrued from these amenities. All three topics are discussed in this subsection.

2.3.5.1 Environmental Amenities

Since Ridker and Henning's (1967) study, the hedonic price model has been a key tool for environmental impact assessment. Smith and Huang (1995) have found 37 empirical studies that related environmental variables to house prices. Some studies concerned the positive effects of environmental amenities, such as the improving house prices resulting from better access to nature (Gibbons et al. 2011; Schaerer, 2007; Cheshire and Sheppard 1995), urban parks (Smith 2010), street trees (Donovan 2010; Morales, 1980) and beaches (Abelson 2013). Contrarily, some studies have examined the negative effects of environmental disamenities, such as water pollution, air pollution, noise and visual intrusions (Dornbusch and Barrager 1973; Lake et al., 1998; Day et al., 2006).

Proximity to nature, parks or views over parks are commonly found as positively influencing property prices, whilst the effects of air and noise pollution are generally determined as negatively influencing property value. Many of these studies used simple spatial separation accessibility to measure exposure to environmental amenities or disamenities. However, pollution and access to nature are also influenced by both the geometry of the built form and the visibility of these amenities (Penn and Croxford 1997). As a result, there is a need to consider how urban form can impacts access to these types of amenities.

2.3.5.2 Urban Amenities

House prices are also influenced by access to urban amenities; one such amenity is retail attraction. The accessibility of retail amenity in hedonic price model is well documented and can be a key determinant of property values (Des Rosiers et al. 2000; Theriault et al. 2005). Sirpal (1994), for example, has studied properties located near shopping centres in Florida. The results showed that properties located close to the shopping centre had higher house prices than properties further away. This effect was positively related to the size of the shopping centre. Rosiers et al. (1996) conducted a similar study for Canadian cities and concluded that a property near a shopping centre has a 5% premium over a similar property further from a shopping centre. In a different context, Addae-Dapaah (2010) found that, in Singapore, housing estates with a shopping centre command higher prices than those estates without one. Extending from previous research, Theriault et al. (2005) have found that the effects of shops on house prices differ statistically according to both the types of shops, trip purposes and household profiles. These consistent results show that retail amenities have a complex and significantly positive effect on house prices. In terms of specification, the majority of early research on access to retail amenity have used simple accessibility measures such as spatial separation and cumulative opportunities. Recent research has begun to examine perceptive accessibility measures which combines commuting data, demographic data and transport data (Theriault et al. 2005). Other urban amenities, including churches (Carroll et al., 1996), cultural assets (Moro et al., 2002) and community centres, also have a similar positive impact on house prices.

The second subtopic in urban amenity is the effect of neighbourhood prestige, as termed in Kain and Quigley's (1970) article. Neighbourhood prestige can be understood as an intangible good made up of a combination of factors, such as school quality, safety and social demographic factors. The effects of school quality are well researched in the UK context, where the influence of pupil spending, pupil-teacher ratio, test scores, teacher salary and tenure, percentage of teachers with an advanced degree and average teacher experience were all found to have significant effects on house prices (Oates, 1969; Black, 1999; Gibbons and Machin, 2003; Gibbons et al., 2009; Gibbons et al., 2011). In 2003, Gibbons and Machin investigated school performance between 1996 and 1999 in the UK. They estimated that school quality has a positive effect on house prices, in which a 1% increase in children's test scores led to a 0.67% increase in house prices.

Another urban amenity related to neighbourhood prestige is safety. The perception of crime and the actual costs of crime have both direct and indirect influences on house prices. These influences were made apparent in a study on Florida by Lunch and Rasmussen (2001), which has found that a 1% increase in crime rate resulted in a 0.05% reduction in house prices. Gibbons (2004) has found that, in London, a one-tenth standard deviation increase in reported incidents of property damage (such as graffiti and vandalism) amounted to a 0.94% reduction in the house prices. One explanation for this decrease is that the perception of crime, evinced by vandalised property, has an indirect effect on the neighbourhood.

The last topic to be discussed is the socio-demographic variable (Watkins 2001). Socio-economic topics, such as ethnicity, have been found to have significant effects on house prices (Zabel 2008; Schafer, 1979; Kain and Quigley, 1970; Lapham, 1971). Zabel (2008) identified four reasons concerning why ethnicity might influence house prices, including prejudice against ethnic minorities, the coupling effect of income and ethnic minorities, discrimination against ethnic minorities obtaining property, and the market expectation concerning where other ethnic groups migrate to. The effects of ethnicity are therefore complex and location-specific and are commonly modelled as the neighbourhood and housing submarket to which the property sits in.

2.3.5.3 Neighbourhood Effect

An amenity influences house prices not only through the amenity itself but also due to the spillover from adjacent properties. This adjacency effect is also known in housing studies as the neighbourhood effect. The idea of the neighbourhood effect was inspired by Tobler's first law of geography, which states that nearer things are more similar than distant things (Tobler, 1970). A large volume of literature has identified the existence of the neighbourhood effect on house prices (Can, 1990). To illustrate this effect, suppose all of the properties in a neighbourhood undergo a façade improvement. This improvement raises house prices through two mechanisms: first, through the direct effect of the property's façade improvement, and second, through the multiple spillover effects from the adjacent property's façade improvement. Contrastingly, the vandalism of all of the properties along the street would reduce the house prices through these same mechanisms. A neighbourhood fixed in decline is characterised by a lack of security, which leads to the neighbourhood gaining a poor reputation, so people and employers leave (Gibbons and Machin, 2008). This indirect effect is also referred to as the spatial autocorrelation effect on house price.

2.3.5.4 Neighbourhood variable limitation

Several limitations remain concerning neighbourhood variables in hedonic price models. One of the major limitation concerns the direction of causality. For example, school quality has been found to significantly correlate with house prices; homes have higher real estate value if they are within the proximity of an excellent school. However, it isn't clear whether the quality of the school influences house prices or if individuals living in wealthy areas influence the quality of the school and thus influence house prices. Same can be said with issues concerning safety and prestige. It isn't entirely clear whether the prestige is influencing house prices or if individuals living in wealthy areas influence the prestige of the neighbourhood and thus influence house price. Understandably, the causal direction is likely to be bi-directional and simultaneous.

A second limitation is the lack of research on how spatial configuration influences both access to the neighbourhood amenity and the neighbourhood effect. The majority of the existing literature focusses on modelling neighbourhood variables using geographic distances. However, both the neighbourhood amenity and effect are influenced by geographic distance and geometric factors of the built environment. For example, aesthetic improvements and vandalism are likely to affect property along

the same street more than the property around the corner; further, they are likely to affect the properties within the same neighbourhood more than those in another neighbourhood. As spatial configuration influences, how pedestrians navigate and identify a neighbourhood, both factors can be affected by the urban environment's spatial configuration.

A third limitation is the lack of research on how urban design influences house prices. Nase et al. (2013) have found that urban design features such as frontage continuity, variety, materiality and massing add to the prices of retail zone A. This conclusion is sensible, as a more aesthetically appealing urban environment will attract more shoppers to the area, holding all other factors constant. However, research in this domain is sparse where several urban design variables have still not been considered in this literature. These variables include the presence of active frontage and the scenicness of the streets.

2.3.6 The Housing Submarket

A key assumption in the hedonic price model is that under equilibrium, the implicit price of the attribute will equal to its marginal willingness to pay (Freeman 1979). However, market equilibrium is not expected to equalise across property markets instantly where attribute price and demand is stationary across space. This inefficiency or market disequilibria can be caused by information asymmetry but also from the differences in demand across socio-economic groups. These differences in implicit prices across local markets have given rise to the concept of housing submarket, which is one of the most discussed topics in housing studies (Maclennan and Tu, 1996; Watkins 2001). Housing submarket is defined as properties that are reasonable substitutes for one another, but poor substitutes for properties in other submarkets (Grisby et al., 1987). Empirically, housing sub-market can be defined as a geographical area where the implicit price of an attribute is relatively constant (Schnare and Struyk, 1976). This sub-field is a major topic in housing research where there is much consensus concerning its existence but little consensus on its definition (Dale-Johnson, 1982; Bourassa et al., 1999). Ignoring the processes that influence housing submarkets will introduce errors in the hedonic price model leading effectively to poor housing policies (Adair et al., 1996). From a real estate perspective, a better understanding of housing submarkets can improve the ability to forecast these markets, reduce investment risk, create more accurate housing policies and improve property valuation (Goodman and Thibodeau 2007; Bourassa, 2002). The next section looks at three empirical topics related to this; namely, the hierarchical nature of housing submarket, the local area geography for defining housing submarkets, and the methodology in defining housing submarkets.

2.3.6.1 Hierarchical nature of housing submarkets

Maclennan, one of the first authors to categorise the different types of housing submarkets, defined three categories of housing submarkets. The first category comprised an entire nation, region or state (Linneman, 1981). The second category included metropolitan areas (Malpezzi et al., 1980). The third category examined submarkets at the metropolitan level, which were segmented by location (centre or suburb), housing quality, race, or income (Straszheim, 1975; Maclennan and Tu, 1996). An important

notion from this research is that housing submarkets are both multi-dimensional and hierarchically nested. To illustrate this concept of hierarchy, we use the following quote from Goodman and Thibodeau (1998):

‘... We consider the value of the house, nested within a neighbourhood, within a school district, within a metropolitan area. Some of these effects may be nested hierarchically, such as blocks within neighbourhoods ...’
(Goodman & Thibodeau, 1998)

Goodman and Thibodeau (1998), in their research, used the Dallas housing market as a case study to demonstrate the existence of a hierarchy in housing submarkets. Orford's (1999) modelled this hierarchical nature, through a multi-level hedonic price model, in which house price variations were deconstructed into variations across enumeration districts, communities and individual properties for the city of Cardiff. One topic related to the hierarchical nature of the housing market is the local area definition. This topic is discussed in the next subsection.

2.3.6.2 Local Area Definition

The second topic concerns the definition of local areas. There is a long history of the research of local areas or neighbourhoods, which ultimately facilitated the creation of terms that are used interchangeably in the housing market literature. The concept of locality or neighbourhood is complex and fuzzy, full of idiosyncrasies; it encompasses spatial, historical, socio-economic and perceptual inner characteristics that change and overlap according to the geographical scale and point in time (Lebel et al., 2007; Galster, 2001; Kearns and Parkinson, 2001). According to Lynch (1967), a city district is an area of homogeneous character (physical, social or functional) that can be further divided into subareas or embedded into larger regions. The more these characters, people, continuity and environment overlap, the more unified the district becomes. The term ‘local area’ is used in this research to represent a geography that is larger than a property but smaller than a district. This concept is not to be confused with the multi-dimensional term ‘neighbourhood’, which often conveys a social meaning. The definition of local area is an important topic in hedonic price studies for estimating local area effects in hedonic price models (Orford, 2001). However, census tracts, which are often used as local area units, have been criticised for their arbitrary definition and inconsistent results (Orford, 1999; Goodman, 1978; Leisham, 2009). Goodman's earlier studies (1978; 1982) have examined the coefficient differences between the block level and the census tract level when estimating a hedonic price model. Goodman (1985) found that segregation indices differed when different levels of aggregation were used. He suggested that this could be attributed to the ‘fuzziness’ and ‘arbitrariness’ of how ‘census tracts’ and ‘block level’ were defined.

2.3.6.3 Submarket Identification

The third topic concerns the identification of housing submarkets, where various researches have dealt with the methods and variables to use for the delineation (Dale-Johnson, 1982; Bourassa et al., 1999; Straszheim, 1973; Schnare and Struyk, 1976). The most common submarket classification is through supply and demand factors. This can include structural factors such as, for example, whether the

property is a house or a flat (Adair et al., 1996; Allen et al., 1995). This can also include spatial factors such as, for example, whether the property is located within, for instance, a school catchment (Galster, 1987; Schnare, 1980) or whether the property is located within a socio-economic segment (Schnare and Struyk 1976; Palm 1978). More often than not, both spatial and structural dimensions are clustered together to form distinct spatial-structural housing submarkets. For example, Watkins (2001) has used different combinations of spatial (north-south-east-west) and structural (dwelling type) factors to construct seven housing submarkets in Glasgow. Adair et al. (1996) have used different combinations of spatial (zones) and structural (dwelling type) factors to construct nine housing submarkets in Belfast. Due to advances in computation and data availability, recent studies have used statistical methods in defining data-driven, spatial-structural housing submarkets. The rationale is that housing submarkets are not simply the construct of a building type or an income group or a school catchment area, but rather that buyers in a specific local market are seeking a combination of these attributes simultaneously. Statistical classification methods such as k-means clustering, principal component analysis (PCA), hierarchical linear clustering and machine learning methods have often been used to define housing submarkets. For example, Maclennan and Tu (1996) and Bourassa et al. (1999) applied PCA and k-means cluster analysis to define housing submarkets in Glasgow, Sydney and Melbourne. Day et al. (2002) used PCA, k-means and then hierarchical clustering to define housing submarkets in Glasgow. The Greater London Authority (2004) employed k-means clustering techniques on standard housing characteristics for defining five to six distinct housing submarkets in London. More recent studies have used commuting patterns for deriving residential sphere submarkets (Park 2013) and fuzzy logic for defining overlapping housing submarkets (Helbich 2015). Some researchers have even adopted the use of expert knowledge for identifying housing submarkets; this achieved more similar results than other, more complicated models (Michaels and Smith, 1990; Bourassa et al., 2002). Due to the complexity of the formation of housing submarkets, it is unlikely that there is only one way to segment a housing market.

2.3.6.4 Limitations in housing submarket research

To summarise, there is an extensive list of research concerning the topic of housing submarket from defining its existence (Schnare and Struyk 1976) to the types of housing submarket (Watkins 2001). This review identified three research gaps in the housing submarket literature. The first topic that is rarely explored in the submarket literature is how different local area units affect housing submarket formation (Orford 2001). The majority of the existing housing studies focuses on using census tracts as the base areal unit to construct housing submarkets. However, the use of arbitrary census tract local areas has been found to produced inconsistent results (Bourassa et al. 1999). Therefore, there is a need for more research in defining local areas and in how local areas influence housing submarket formation. This thesis investigates this topic in the third empirical chapter.

The second topic that is rarely explored in the submarket literature is the inclusion of spatial configuration parameters in housing submarket formation. Xiao et al. (2016) recent study has provided a template to including centrality indicators in delineating housing submarkets. The study

found that market segmented by street morphology corresponds to ones defined by building type. The authors stressed the advantage of using spatial network methods in avoiding the use of arbitrary geographic boundaries. This research will also extend on this notion by using similarly the information of a spatial dual graph in the formation of housing submarkets.

The third topic that is rarely explored in the housing submarket literature is the lack of research on the spatial temporal stability of housing submarket (Jones et al.2003). Bourassa et al. (1999) for example stressed the need to test the stability of housing submarket boundaries. This is important as instability of housing submarket can lead to poor policy allocation. The modelling of such notion is challenging methodologically. Under this notion, Helbich (2015), explored the notion of identifying fuzzy housing submarket for the city of Vienna using a non-linear hedonic pricing model. Jones et al. (2003), on the other hand, constructed repeat-sales indices in Glasgow to examine the stability of housing submarkets between 1984 and 1997. Resonating with previous research, Jones study found that results from cross section studies of housing submarket can be misleading. However, the authors also found that there is stability in housing submarkets throughout this study period.

2.3.7 Other Controlled Variables

As noted in previous meta-studies, factors pertaining to the market, individual finance, and land ownership also influence house prices. The property's duration on the market, the property's tenure (freehold or leasehold) and the individual's credit rating all have effects on house prices. A leasehold property, on average, is less valuable than a freehold property (holding all other variables constant). This effect is due to both the length of the contract and the fact that leaseholder requires to pay an additional rent to the freeholder for owning a leasehold of the property. As this thesis focuses on how urban built form influences house prices, individual financial factors are not discussed further in the review. Instead these variables are considered as controlled variables in the study.

2.3.8 Inter-city Regional House Price Variables

This section briefly discusses three empirical topics that relate to inter-city regional house price variables, including wage and productivity, supply constraint, policies and investment factors. These factors are discussed briefly here as they are not the focus of this research.

2.3.8.1 Wage, Productivity and Population

Economic fundamentals suggest that differences in inter-city house prices can be attributed to differences in wages, population and productivity between cities. The Rosen-Roback's model translated this into a wage-location trade-off problem or a hedonic model of the inter-city location. Glaeser (2008) has suggested that the differences in the productivity between cities can be explained by population variances. These differences are due to agglomeration benefits, such as access to customers, resources and new ideas, labour specialisation and spillover effects (Marshall, 1890; Jacobs, 1961; Krugman, 1991). These agglomeration effects are also evident in the works of

Bettencourt (2012) where productivity was found to increase disproportionately to population thereby areas with a greater population, brings greater relative productivity, wages and thus higher house prices.

2.3.8.2 Supply-side, Physical Constraint and Construction Costs

Economic fundamentals also suggest that differences in inter-city house prices can be explained by supply constraints and construction cost differentials (Glaeser et al., 2008). Housing supply is a primary factor in influencing house prices and can come in both physical and regulatory forms. Physical factors, such as the topography and the waterfront, and regulatory factors, such as land use policies, put a constraint on where houses can be built. Hilber and Vermeulen (2010) and Glaeser et al. (2010) have noted that places with limited housing supply are more sensitive to demand shocks, which leads to higher house prices (Glaeser et al., 2008). However, the supply effect is primarily an inter-city effect as opposed to an intra-city issue. For their case study of Boston, Glaeser and Ward (2009) found that a restriction in the housing supply in one neighbourhood did not necessarily raise the other house prices in that neighbourhood. This result was largely due to the abundance of substitutes in similar neighbourhoods within the same metropolitan area. This conclusion does not suggest that the local supply effect is non-existent, but rather that these effects are weaker within large metropolitan regions. In addition, inter-city house prices are also influenced by construction costs. Gyourko and Saiz (2006) found that one-fifth of inter-metropolitan house price differences in the US could be explained by construction cost differentials. Places with lower construction costs were associated with lower house prices and vice versa. This notion can be used to explain the urban decline in the US Rust Belt, where house prices are below construction costs (Glaeser and Gyourko, 2005). Again, these effects occur primarily between cities, as construction costs are relatively equal within cities.

2.3.8.3 Policy and Investment Factors

External forces such as government policies and foreign direct investments can also have a strong effect on inter-city house price variations. According to economic fundamentals, fiscal instruments, such as the Bank of England base rates, are important drivers that boost the demand and prices in the market as lower interest rates can bring cheaper mortgages (Himmelberg et al., 2005). However, the extent on how these factors influence house price is more complex than what traditional economic fundamentals would suggest. A noticeable example is recent works from Glaeser et al. (2009) whom have shown that credit access had only a minor effect in the price shifts during the housing crisis in 2011. Foreign direct investment is another important factor that influences house prices in London, especially in the prime central area. The reasons for this are multiple and can be attributed to a combination of the cheaper pound, relative political and financial stability in the UK and investor needs to spread risks. The way in which housing demand factors can influence house prices is significantly different for this target market than others. To fully understand the implications of these inter-city house price factors is beyond the scope of this research. As a result, these factors are only briefly discussed here.

2.3.9 Limitations in empirical studies using the hedonic price approach

In summary, there are research gaps in empirical studies using the hedonic price approach. First, there is a lack of consideration regarding how spatial configuration influences house prices across multiple scales and variables. This limitation can be related to the configuration of a home at the property scale, the street network formation at the neighbourhood scale and the transport system configuration at the city scale. For example, geometric factors such as the spatial configuration of the street network, which influences how pedestrians navigate and identify a neighbourhood (Hillier and Iida 2005; Law et al. 2015), have not been considered in hedonic price modelling fully.

The second limitation concerning the use of the hedonic price model involves the direction of causality. Much of the econometric research has focussed on the specification design in retrieving the causal effects of such outcomes. For example, Black (1999) has used the novel boundary discontinuity model to estimate the effects of school quality using school catchment boundaries. Gibbons and Machin (2005) used difference in differences models to estimate the effects of transport innovation using house price panel data. Despite efforts to unveil the potential causal effect using econometric design, the causal mechanism remains unclear. This uncertainty is a general limitation to the use of the hedonic price regression model as an empirical approach in highly complex domains, such as social sciences, where a controlled clinical trial is not feasible. The next section will briefly discuss the empirical specifications of the hedonic price regression model.

2.4 Empirical Specification

The hedonic price model has traditionally been specified as a regression model, where the house price variances are explained by a series of independent variables. The basic form uses an ordinary least squares (OLS) estimator. As is common with regression methods, these models are subject to statistical problems, such as omitted variable bias, heteroscedasticity, confounding variables and spatial autocorrelation. As a result, new types of regression models have been developed. Variation to the original model includes the use of an Artificial neural network model (Kauko, 2002), a spatial hedonic regression model (Pace et al., 1998), a Geographic Weighted Regression model (Fotheringham, 1999), a multilevel hedonic regression model (Orford, 1999) and a panel data fixed effect regression model (Gibbons and Machin, 2005). Figure 2.9 below describes this process. This section begins by looking at the OLS multiple variable regression model. It then examines common statistical biases and concludes by discussing three of these advanced regression models. As this thesis focus is architectural and not on the econometric/statistical specification, only standard regression techniques are reported. More advanced machine learning techniques such as the use of artificial neural network regression model and deep learning neural network frameworks are not reported and considered for this research (Kauko et al., 2002).

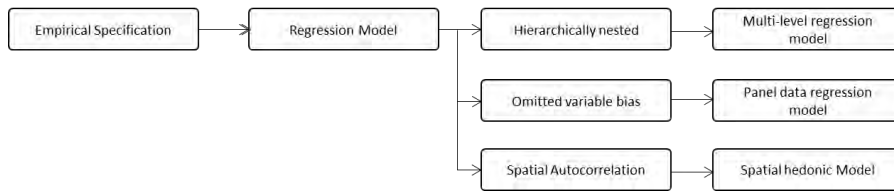


Figure 2.6 Empirical specification for the hedonic price regression model.

2.4.1 Multiple Variable Regression Model

The multiple variable regression model is an additive linear regression model in which the dependent variable is regressed against a set of independent variables by minimising the sum of squared errors. The functional form of the regression often transforms either the dependent or independent variable. The most common is the semi-log form (Equation 2.4), where the logarithm of the dependent variable Y is regressed against a vector of independent variables X , or the log-log form (Equation 2.5), where the logarithm of the dependent variable is regressed against a vector of the logarithm of the independent variables. Another type of transformation commonly used in hedonic models is the Box-Cox transformation.

$$\text{Log}Y = \beta X + u$$

Equation 2.4

$$\text{Log}Y = \beta \text{Log}X + u$$

Equation 2.5

2.4.2 Regression Model Limitations

Well-known empirical topics in hedonic price regression models include multicollinearity, heteroscedasticity, spatial autocorrelation and omitted variable bias. In order to account for these statistical problems, a robust statistical design is required.

2.4.2.1 Multicollinearity

The term multicollinearity refers to the correlation between the independent variables in a multiple variable regression model. Multi-collinearity is a major topic in hedonic price studies as variables are understandably path dependent. Estimates can become inflated or deflated when collinearity is present. An obvious example is that the age of the building will covary negatively with distance to CBD due to the outward growth process of a city from the centre. However, independent variables are bound to correlate; the primary question concerns the extent to which this correlation will occur. The variance inflation factor (VIF) is the most common method for measuring the extent of this multicollinearity. The VIF calculates how much the variance of an estimated regression coefficient has increased after each independent variable X_i has been correlated with the other independent variables.

$$X_i = \beta X_j + u$$

Equation 2.6

The VIF is then calculated as follows:

$$VIF = \frac{1}{1 - R_i^2}$$

Equation 2.7

Multicollinearity becomes a problem when $VIF > 10$. If multicollinearity is detected, one can choose to retain the variable due to theoretical reasoning, construct a more parsimonious model or combine the variables through principal component analysis.

2.4.2.2 Heteroscedasticity

The term heteroscedasticity refers to the error term's unequal variance. The OLS estimate remains unbiased but rather inefficient because the true variance and covariance are not correctly estimated. This poor estimation leads to inaccurate standard errors, which can influence the validity of the statistical inferences. The most common method for identifying heteroscedasticity is plotting the errors against each independent variable to visually determine whether an association exists. If heteroscedasticity is detected, the original regression can be corrected by transforming the independent and dependent variables or estimating robust standard errors.

2.4.2.3 Omitted Variable Bias

One of the most common statistical problems in an OLS regression model is the omitted variable bias. The presence of the omitted variable bias violates the Gauss-Markov theorem assumption, where the error term is correlated with both the independent and dependent variables. This can lead to estimates being misevaluated. Variations of the hedonic regression model have been developed to account for these errors.

2.4.2.4 Spatial Autocorrelation

Spatial effect or spatial autocorrelation is commonly defined as the spatial association between adjacent observations of the same phenomenon (Anselin, 1988; Pace et al., 1998). In this context, positive spatial autocorrelation means the price of a property is similar to the price of its neighbours, and negative spatial autocorrelation means the price of a property is dissimilar to its neighbours. Empirically, the existence of spatial effect violates one of the key OLS assumptions, where data are assumed to be independent of each other.

As a result, a number of indices for measuring spatial autocorrelation have been developed. The most popular are Moran's I Index, which was developed by Moran (1950), and the local variation, developed by Anselin (1995). Moran's I is a global index that correlates a dwelling's sold price with a neighbouring sold price and can be calculated more formally, where w is the weight matrix, X is the price of the observation, \bar{X} is the mean price, and N is the number of observations.

$$I = \left(\frac{N}{\sum \sum w_{ij}} \right) * \frac{\left(\sum \sum w_{ij} (X_i - \bar{X})(X_j - \bar{X}) \right)}{\sum (X_i - \bar{X})^2}$$

Equation 2.8: Moran's I equation (Goodchild, 1986; Anselin, 1988).

The results range from -1 (indicating perfect dispersion) to +1 (indicating perfect correlation) and 0 (indicating a random pattern). Positive spatial autocorrelation means that the house price is similar to its neighbour's, which suggests strong homogeneity. A large body of literature has consistently shown observable spatial effects on house prices across space and time, where the distribution is highly clustered and not random. However, failing to incorporate the spatial effects into the regression model will result in bias and misleading coefficients (Anselin, 1988). Since the identification of the spatial effects, a class of regression models have been developed, which are known to account for such effects. This is discussed in the next section.

2.4.3 Advanced Regression Models

To account for the different problems of the OLS regression model, advanced regression models have been proposed. These models include the panel data fixed effect regression model, which accounts for omitted variable bias, the spatial hedonic model to account for spatial effects, and the multilevel regression model to account for both neighbourhood and submarket effects.

2.4.4 Spatial Regression Models

Extensive research has been conducted concerning spatial autocorrelation effects (Can, 1990; Anselin, 1988). The most common approaches to account for these spatial effects are the spatial lag/error models also known as spatial autoregressive model (SAR) and the geographically weighted regression model (GWR). The spatial lag model assumes autocorrelation in the response variables, while the spatial error model assumes autocorrelation in the error term. As a result, a set of weights is given to either the lag or the error term in the regression model. The next common approach to account for the spatial effects is the GWR (Fotheringham et al., 2002). This method is essentially a non-parametric local regression model that assumes a separate regression model is estimated for each data point. The rationale behind the GWR is that parameters vary across different parts of the city where people's tastes and attitudes differ geographically. This technique is widely used in the field of regional science. These methods have been applied to environmental quality research (Kim et al., 2003), neighbourhood effect research (Tse, 2002), accessibility research (Osland, 2007) and land use and transport modelling research (Lochl, 2010).

2.4.5 Panel Data Fixed Effect Regression Models

A central concern in a typical OLS hedonic price regression model is that important factors are incorrectly omitted from the model. This omission can include qualitative factors, such as urban design quality or the perception of crime, which are often difficult to measure. When the unobserved time-invariant characteristic is correlated with both the dependent and the independent variables, this can lead to biased estimates. One method to account for the unobserved variance is the application of the panel data fixed effect regression model, which examines within-property differences. By looking at the within-property differences, the property's unobserved time-invariant characteristics can be dropped. More simply, if an omitted variable did not change over time, then any change in price cannot be caused by the omitted variable but rather by the time-variant characteristic. This technique is used for the first analytical chapter and will be discussed more formally there. A related method is the difference-in-difference model which will also be discussed in the analytical chapter.

2.4.6 Multilevel Regression Models

Another advance in the regression technique to construct a hedonic price model is the multilevel regression model (Goldstein, 1987). This method accounts for both the neighbourhood effect and the submarket effects in a hedonic price model (Orford, 1999; Goodman, 1998). Examples include Orford's (1999) house price study in Cardiff. He found (significantly) that house price variations from the grand mean could be deconstructed into variations across districts, communities and individual properties. This technique is used in the second and third analytical chapters.

2.5 Discussion

Research examining intra-city house price variations often focuses on estimating the marginal implicit price for an amenity using the hedonic price approach (Rosen, 1974; Cheshire and Sheppard, 1998). The concepts of spatial equilibrium, hedonic price theory, and housing submarkets provide the foundation for the approach, and a large volume of research has been conducted that deals with the theoretical, methodological, and empirical concerns. Within the field of environmental and real estate economics, this method provides a robust and established framework in estimating the economic value of an intangible good. Within the field of geographical science, the hedonic price model's predictions can be used as inputs in land use models. This process can reduce investment risks, provide better housing policies and help architects and planners make more informed decisions when designing and planning neighbourhoods.

Several limitations have been highlighted in the past by scholars. The economic assumption of market equilibrium (i.e. an individual who buys and trades properties in markets with perfect information) is unrealistic which can lead to an inequality between the implicit price of an attribute and its marginal willingness to pay (Freeman 1979). There are also the general concerns on causal inferences being improbable in social science using econometric methods. Despite these limitations, this assumption does not render the technique invalid. A lot of effort from the econometric literature have been

targeted in developing robust research design. This includes the development of methods such as boundary discontinuity, spatial hedonic model and panel data model. The hedonic price method also has many advantages in retrieving information from observed property characteristics. For example, in the availability of the data and the quantity of the data. This chapter concludes by summarising the research gaps on relating spatial configuration factors and house price in the review.

2.6 Research Gaps - Spatial Configuration

The review has identified three research gaps (depicted in Figure 2.10 in red) to be studied in this research concerning the association between spatial configuration factors and house prices. The first concerns the accessibility effect, the second concerns the local area effect, and the third concerns housing submarket effects. This section summarises these three research gaps.

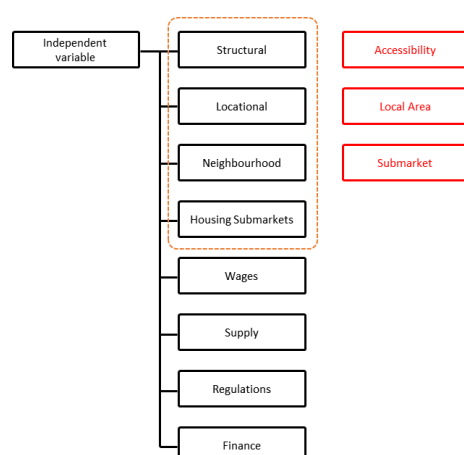


Figure 2.7 The review identified three research gaps concerning the association between spatial configuration factors and house prices.

The theory supporting the use of accessibility derives from the concept of spatial equilibrium and its exposition through a monocentric model. The model operates through a bidding process, whereby the people who capitalise the most from the assets acquire the land rights in a property market. Based on the monocentric model, the location differential is traditionally estimated in the form of Euclidean distance to the CBD in hedonic price modelling (Kain and Quigley, 1970). One limitation is that this requires the endogenous definition of a CBD location; however, inner-city decline, coupled with rising suburban employment, has led to the diminishing influence of central places (Heikkila et al. 1989). This led to the use of multiple employment accessibility models. The motivation behind these approaches is that they move away from the idea that all economic activity is concentrated in a single, dimensionless point and embrace a more heterogeneous distribution of employment. Recent research uses gravity-based accessibility measures, which capture both the size of the employment and its spatial separation simultaneously (Adair et al. 2000; Ahlfeldt, 2010). While this is an improvement from traditional methods, several limitations remain. The first limitation is that geographic accessibility measures require information concerning employment location. The problem is that exact employment information is often

not both geographically and temporally available. For example, the UK census only provides employment data every 10 years in large census tracts. A second limitation is that these geographic accessibility measures focus more on specific accessibility effects and less on general accessibility effects, as explained by Webster (2010). This gap has led to an interest in research within the field of space syntax, which has methods for the quantification and valuation of general accessibility. Empirical research has found significant positive associations between spatial network accessibility and house prices in London (Law et al. 2013). Despite the identification of these associations, there is limited research comparing general (geometric) and specific (geographic) accessibility effects on house prices. Studies that have researched differential distance in associating accessibility and house prices are also limited. This topic is explored in the first analytical chapter.

A little-explored topic within the field of urban planning and housing studies is the definition of the local area unit. In the past, administrative census tracts were often used as local areas to measure neighbourhood effects on house prices. The use of these arbitrary definitions has led to inconsistent results. This inconsistency stems partly from the lack of consideration of the urban built form experienced at the street level in forming a local area. Previous spatial configuration research suggests that the topology of the street network relates not only to how we move in space but also to how we associate with a place (Dalton et al., 2006; Yang and Hillier, 2007). As a result, there is a need to consider spatial configuration in the identification of local areas at the street level and how these new types of local areas can capture the neighbourhood effect on house prices. This topic is explored in the second analytical chapter.

The housing submarket has been one of the most popular topics in housing studies over the past few decades. Despite consensus on its existence, there are general disagreements on the methods and variables for housing submarket identification (Watkins, 2001; Schnare and Struyk 1976; Bourassa et al., 1999). One topic that is rarely discussed is the geography used in constructing housing submarkets. In most housing submarket research, administrative census tracts are used to construct these submarkets. Similarly, using these arbitrary local areas has led to inconsistent results in the hedonic price model. Again, this can stem from the lack of consideration of the urban built form when constructing the housing submarket. There is, consequently, not only a need to consider spatial configuration methods in defining the local area, but also in defining housing submarkets. This topic is explored in the third analytical chapter. The next chapter describes in detail the spatial configuration methods used in the analytical chapters.

Chapter 3

Spatial Configuration Methods

3.1 Introduction

The previous chapter reviewed the hedonic price model and identified several research gaps from a spatial configuration perspective. This chapter focuses on introducing the space syntax theory and the spatial configuration method for this thesis. The chapter is organised into four sections. The first section introduces space syntax and the conceptual framework that links spatial configuration and house prices. The second part describes the method for the first research strand concerning geometric accessibility. The third segment describes the method for the second research strand concerning Street-based-local-area (St-LA). The fourth section describes the method for the third research strand concerning Street-based-housing-submarket (St-HS).

3.1.1 Space Syntax

'Architecture determines to a substantial extent the degree to which we become automatically aware of others, both those who live near and strangers, as a result of living out everyday life in space.'

Hillier and Hanson, 1984

Space syntax is a set of theories and techniques that link space and society. It is based on research by Bill Hillier, Julienne Hanson and their colleagues at the Space Syntax Laboratory, University College London (Hillier and Hanson, 1984). Space syntax suggests that where people are, how they move and how they interact are fundamentally influenced by the geometry and configuration of space. Space syntax views buildings as geometry that orders spatial relations, rather than as objects. Space syntax originates from two fundamental propositions (Hillier and Vaughan, 2007). The first of these is that space is not a background to human activity but is intrinsic to it. The second is that space is first and foremost configurational. Spatial configuration means that what happens in any individual space is fundamentally influenced by the relationships between that space and the network of spaces to which it is connected. It is the understanding of space through simultaneous interdependence. Borrowing from graph theory, this interdependence is translated into spatial networks, which are the fundamental units in space syntax for measuring spatial configuration.

3.1.2 Space Syntax as a Theory for Cities

Space syntax as a theory for cities suggests that spatial relations influence land use via the activity the grid generates. This activity, in turn, changes the configuration of the city through a feedback mechanism. Figure 3.1 shows this conceptualisation, which is adapted from Wegenaar's (1994) conceptual diagram on land-use transport (LUTI) models. The adapted diagram shows that spatial configuration is the driver behind urban functions.

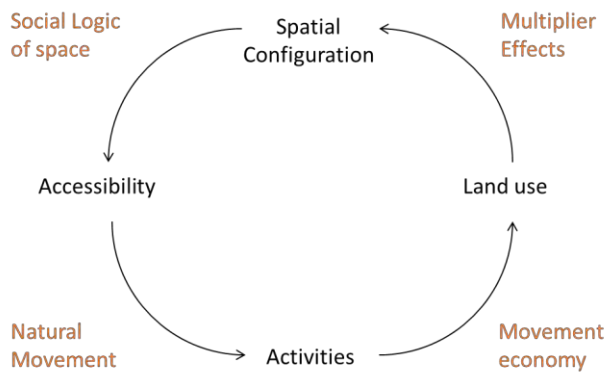


Figure 3.1 Conceptualising space syntax as a theory for city. Diagram has been adapted from Wegenaar's (1994) LUTI diagram.

Empirically, the importance of spatial configuration is shown through the primary association between spatial configuration properties and movement (Penn et al., 1998), land-use distribution (Hillier et al., 1996; Scoppa and Peponis 2015; Ortiz-Chao and Hillier 2007) and density distribution (Law and Versluis 2015). The importance of spatial configuration is also shown through the correlations between spatial configuration and three types of performance. The first is economic performance, which includes office rent (Deysellas, 1997), council tax (Chiaradia et al., 2012) and house prices (Law et al., 2013; 2015). The second is social performance, which includes persistent poverty (Vaughan 2005) and social trajectories in historic cities (Karimi 1998). The third is environmental performance, which includes urban pollution levels (Croxford and Penn 1995; 1998). Bringing all of this together, Hillier (2009) suggested that spatial configuration is intrinsic to these three pillars of sustainability; namely, the social, economic and environmental performance of cities. This research intends to focus on the understanding of the difference between the spatial configuration and the economic performance of cities, where the spatial configuration is not the background to economic performance but intrinsic to it.

3.1.3 Three Research Strands in Linking Spatial Configuration and House Prices

When linked to economics, accessibility inherently becomes a scarce resource produced by the spatial configuration of the city (Narvaez et al. 2012). In a concentric city (Burgess 1975), the spaces in the first ring will always cover less space than the concentric rings around it. This inequality or spatial scarcity results in location differences in housing markets, where accessibility can be implicitly priced (Webster 2010). Simply, some individuals will pay more to live in isolation, while others will pay more to live near other people or in other words agglomeration.

Webster differentiated accessibility into two categories: special accessibility, which concerns access to a specific land use, such as health care or parks, and general accessibility, which includes spatial network accessibility. The greater the scarcity of a particular type of accessibility, the greater the effect it has on house prices. Based on Webster's framework, Xiao et al. (2015) and Law et al. (2013) found significant accessibility effects on house prices using spatial configuration methods. Despite

this evidence, research is still very limited in regards to looking at how ‘accessibility’ or other ‘spatial goods’ produced by spatial configuration relate to the housing market.

This research intends to take the space syntax theoretical argument one step further by suggesting spatial configuration is producing spatial network accessibility effect and influencing the formation of community and housing submarket which in turn impacts upon residential location choice and house price pattern of a city. This conjecture can be drawn from the theory of natural movement and movement economy, where space syntax theory provides a strong argument that spatial configuration is a lead cause for natural movement patterns and land use pattern and in this research, house price patterns. From this conceptual framework, this research suggests linkages between spatial configuration and house prices across three strands of analytical research.

The first research strand suggests that spatial network accessibility, termed here as ‘geometric accessibility’, is associated with house prices. This research argues that geometric accessibility has both a unique effect on house prices and an overlapping effect with geographic accessibility measures. Similarity between the two accessibility measures can be drawn from the theory of the movement economy (Hillier et al 1996) where spatial configuration is a determinant of commercial land use via the “natural movement” it generates. The commercial land uses in turns create “destination-movement” that is more associated with geographic accessibility model. Differences between geometric and geographic accessibility can come from the general accessibility attraction as Webster calls it or the cognitive affordances that spatial configuration produces and are not considered in geographical models. This notion leads to the first hypothesis that geometric accessibility correlates significantly with house prices; this hypothesis is explored in the first analytical chapter.

In existing housing studies, most research uses administrative local areas to estimate the neighbourhood premium on house prices; for example, the premiums to live in Kensington or Crouch End. The second strand of research argues that administrative local areas, due to their arbitrary definition, do not necessarily capture the full neighbourhood effect on house prices. The reasons for this arbitrariness are that individuals experience the environment at the street level and that this connection between streets can also influence place formation (Dalton 2006; Yang and Hillier 2007). Over time, these subtle spatial differences can bring about great socio-economic variances (Schelling 1969). This notion leads to the second hypothesis that St-LAs correlate with house prices, which is explored in the second analytical chapter.

Administrative local areas are also used to construct housing submarkets. The third research strand argues that administrative local areas do not necessarily capture the entire submarket effect on house prices. The reason for this is similar to those stated above: as we experience the urban environment at the street level, the submarket, formed by spatial configuration factors, is potentially more accurate.

A plausible explanation is that buyers whom are attracted to live in a certain housing submarket would attempt to buy in a local area within the submarket or, if they cannot afford it, at the local area connected to it. An example, are those whom wanted to live in Shoreditch but cannot afford to might preferred to move to areas that are connected to Shoreditch such as Hackney or Dalston rather than elsewhere. This preference leads to housing market formed by street-based local area to be potentially more accurate. This notion leads to the third hypothesis that the St-HS correlates with house price, which is explored in the third analytical chapter.

3.1.4 Discussion

The benefits of using spatial network configuration on house price research are many. By focusing on how spatial relations influence activities and functions, employment can be endogenous and not determined a priori. This theoretical proposition produces a powerful tool in which changes in the spatial network can bring about changes to the individual and to society. This method allows architects, urban designers, and planners to influence socio-economic outcomes by influencing space. This application is demonstrated through the projects carried out by Space Syntax Limited (Karimi et al., 2012). Another benefit is that the quantitative evidence can be used to inform policies or taxation or to build more economically equitable and sustainable places.

There are, however, limitations to this approach. One clear theoretical argument is the paradox of space and function. For example, changes in the spatial network can bring alterations to an area, but changes in an area can also cause shifts in the spatial network. One can posit that the actual process is likely to be complex. Consequently, the results can only suggest association rather than causation. Part of this research responds to these limitations by providing more robust evidence through econometric techniques. However, future researchers shall require more robust research designs to model house prices and represent dynamic real-estate market processes through mixed-methods.

To summarise, this section has introduced space syntax as a theory and an analysis technique, which links the spatial configuration of the urban built form with house prices. Despite these limitations, repeated evidence shows a strong basis for the use of spatial network configuration as a method for measuring urban built form and in association with economic performance. This thesis argues that the economic value and the effects of spatial network configuration can be examined across three scales: the accessibility effect, the local area effect, and the submarket effect. Figure 3.2 illustrates these three related research strands. The next section introduces these three research strand methods, which will be used in the analytical chapters.

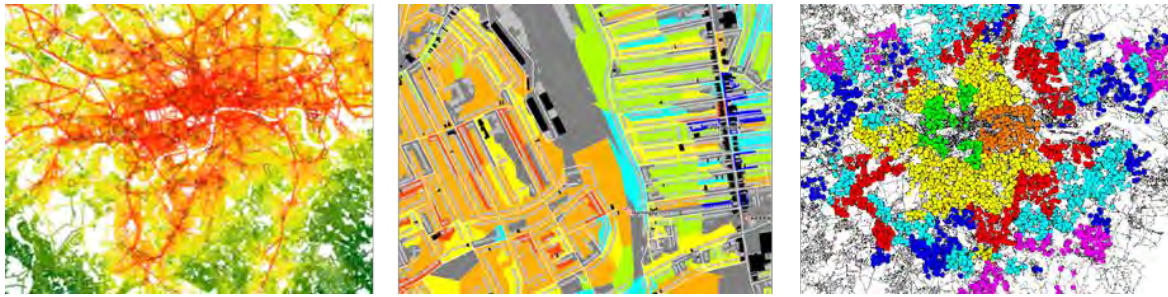


Figure 3.2 The three research strands.

- a. Geometric accessibility.
- b. Street-based location area.
- c. Street-based housing submarkets.

3.2 Research Strand One: Geometric Accessibility

In Hansen's (1949) seminal paper, accessibility is defined as a measure of the potential interactions or the relative proximity or nearness of individuals or places to all others in an environment. Accessibility measures are made up of four inter-related components (Geurs and Wee 2004). These components include a land-use component that deals with the amount of spatial opportunities at a particular location; a transport component that deals with the distance, cost or travel time in accessing the opportunity; a temporal component that deals with the time of the day at which the opportunity is available; and an individual component that deals with how opportunity differs between individuals, demographics and activities. Differences between these four components give rise to the different types of accessibility measures. This research adopts the definition of Jiang et al. (2002), who classify accessibility into two categories: geographic accessibility, which concerns the attraction and distance between places, and geometric accessibility, which concerns the spatial network itself.

The first research strand argues that geometric accessibility correlates with house prices. The reason for this is that geometric accessibility can capture attraction effects, which are generated from spatial network configuration. Figure 3.3 illustrates this comparison. On the left is geometric accessibility, which measures access to every space in capturing the spatial network effect. On the right is geographic accessibility, which measures access to specific spatial opportunities, such as employment opportunities.

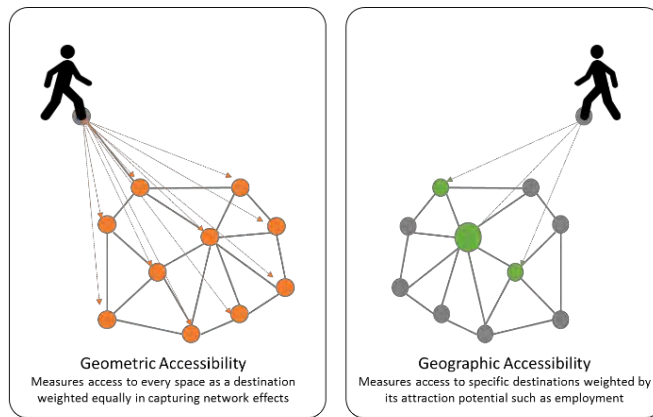


Figure 3.3 Geometric accessibility considers the network effect and geographic accessibility considers attraction and distance effect.

The following section, mapped in Figure 3.4, illustrates the method used to calculate these two types of accessibility measures. In order to study the association between accessibility and house price, we first describe the network representation and then the accessibility measures to be calculated.

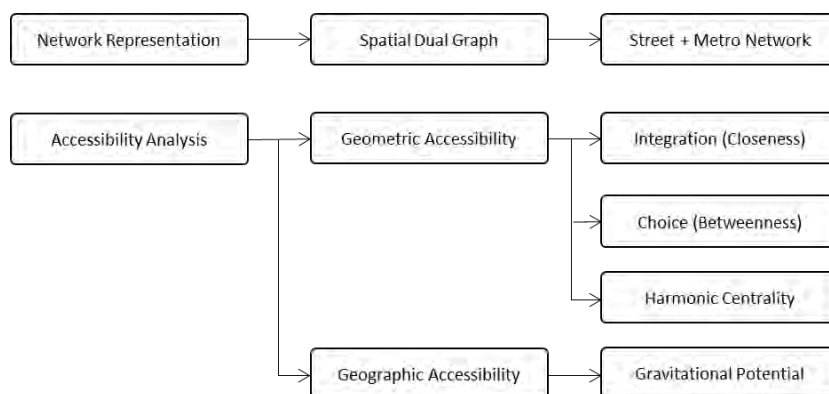


Figure 3.4 Accessibility analysis explanation for the first research strand.

3.2.1 Spatial Network Model

In space syntax, there are multiple ways to represent cities. Two of the most common spatial units are the axial line model and the segment line model. In an axial line model, each space is drawn to represent the longest line of sight between all connected convex spaces. In a segment line model, the segment is the section of axial line or street lying between two intersections. This research selects the segment line model produced from the OS Meridian 2 (Ordnance Survey 2015) road centre line dataset, with the manual addition of pedestrian paths. This spatial network model is used in all three research strands. The first research strand also includes the London Underground network in calculating accessibility. The main reason for this inclusion is in a large agglomeration such as London, a significant proportion of pedestrians uses the public transport network to get to different destinations. The second reason for this inclusion is it correlates better to other economic variables such as passenger flows (Law et al. 2013). An obvious limitation is the exclusion of the rail, bus and ferry networks; this drawback is discussed in the analytical chapter.

3.2.2 Spatial Dual Graph Representation

To calculate accessibility measures, the spatial network needs to be translated into a graph made up of a set of nodes **V** and edges **E**. Spatial graph representation can come in two forms, namely the primal graph **PG** and the dual of the primal graph **DG** (Batty, 2004; Porta et al., 2006). Figure 3.5 illustrates these two types of graphs.

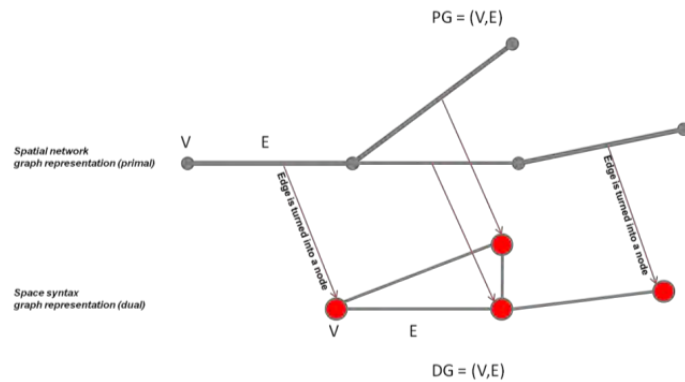


Figure 3.5 The spatial network graph definition.

a. Primal graph representation at the top.

b. Dual graph representation at the bottom.

In a primal graph **PG**, streets are edges **E** and junctions are nodes **V**:

$$PG(V, E)$$

Where

v is the set of nodes (junctions)

e is the set of edges (street segments)

Equation 3.1

In a dual graph **DG**, streets are nodes **V** and junctions are edges **E**. One of space syntax's key contributions to urban design is the translation of cities' spatial networks into spatial planar dual graphs. This research uses this representation, as it allows the accessibility to be measured on the street that the property faces. The next section defines the accessibility measures used for the study.

$$DG(V, E)$$

Where

v is the set of nodes (street segments)

e is the set of edges (junctions)

Equation 3.2

3.2.3 Accessibility Definition

Accessibility is defined as a measure of the potential interactions or the relative proximity or nearness of individuals or places to all others in an environment (Hansen, 1949). The most common of these include spatial separation, which measures the minimum distance to a particular point or attraction (Equation 3.3); cumulative opportunity, which calculates the sum of access to spatial opportunities at a particular radius (Equation 3.4); gravity-based measures, which determine accessibility in proportion to both the size of the attraction and the distance between attractions; time-space measures, which are individual-based measures of accessibility, according to the time of the activity and individual differences; utility measures, which monetarise accessibility benefits; and lastly, network-based measures, which determine access to spaces rather than places (Bhat, 2000; Curtis and Scheurer, 2007).

$$Sep_{ij} = \min d_{ij}$$

d_{ij} is the minimum distance(impedance) between the origin i and the destination j

Equation 3.3

$$CO_{it} = \sum o_{jt}$$

o_{jt} is the number of attractions j that can be reached within the radius t

Equation 3.4

Two of these types—namely, gravity-based measures and network-based measures—are discussed in the next section as geographic and geometric accessibility measures; these measures are applied in the first analytical chapter.

3.2.3.1 Geographic and Geometric Accessibility Methods

Jiang, Claremont, and Batty (2002) defined geographical accessibility as a function based on its attraction and the impedance between the origin and the destination. These measures are defined as:

$$GA = f(W, D)$$

Where

GA is the geographic accessibility measure

W is the number of spatial opportunities or an index of attraction

D is the measure of impedance

Equation 3.5

One form of such a measure, analogous to Newtonian physics, is gravitational potential (Hansen, 1949). Accessibility, in this instance, is a function that is positively proportional to attractions and inversely proportional to the distance between the household location and employment. This is summed up for each employment region or building. The distance function usually takes an exponential form, but this research uses a linear form to allow distance to be compared with closeness and harmonic centrality measures in the analytical chapters.

$$gp_i = \sum_j o_j d_{ij}^{-1}$$

Where

gp_i is the gravitational potential

o_j is the attraction at j

d_{ij} is a measure of impedance between i and j

Equation 3.6

Jiang et al. (2002) noted that, when the focus of the measure is on the spatial network itself rather than on places, these measures are defined as geometric accessibility. This type of measures can be expressed as equation 3.7, where the attraction is the street itself:

$$GM = f(N, D)$$

Where

GM is the geometric accessibility

N is the street network attraction

D is the measure of impedance

Equation 3.7

A popular form of geometric accessibility is space syntax integration or closeness centrality in graph theory (Hillier and Hanson, 1984; Hillier et al., 2012). Space syntax integration measures the reciprocal average shortest path between every origin *i* to every destination *j*, or more simply, the to-movement potential in the system (Sabidussi, 1966; Freeman, 1977; Hillier and Iida, 2005). The nominator is the total number of nodes reach from i to give the following:

$$cc_i = \frac{N_i}{\sum d_{ij}}$$

Where

cc_i is the measure of closeness centrality

N = ∑n is the total number of nodes reachable from i

d_{ij} is the measure of the impedance between i and j

Equation 3.8

A second geometric accessibility measure introduced in this study is harmonic centrality (Boldi et al., 2014), which applies an inverse distance function to reach each node, similar to the gravitational measure. The measures do not account for employment differentials but rather node count differentials. Therefore, being situated closer to more connected nodes brings higher accessibility than being further away.

$$hc_i = \sum \frac{1}{d_{ij}}$$

Where

hc_i is a measure of the harmonic centrality at i

d_{ij} is the measure of the impedance between i and j

Equation 3.9

These sets of measures, namely, gravitational potential, harmonic centrality, and closeness centrality, capture the positive accessibility effects of location in a hedonic price model. A different class of geometric measures is the space syntax choice or betweenness centrality in graph theory (Hillier and Iida, 2005). Rather than measuring how accessible or central a street segment is relative to other segments, this measure captures the through-movement potential. Betweenness centrality measures the sum of the shortest path (θ) overlap for a particular segment i between all pairs of origins s and destinations t . Simply, betweenness centrality captures the through-movement potential of a street segment (Freeman, 1977; Hillier and Iida, 2005). Betweenness centrality, a graph-based measure associated with pedestrian flow, is represented by the following equation:

$$bc_i = \sum_{s \neq i \neq t} \frac{\theta_i(st)}{\theta(st)}$$

Where

bc_i is the measure of betweenness centrality at i

$\theta(st)$ is all shortest path between s and t

$\theta_i(st)$ is all the shortest path between s and t that overlaps at i

Equation 3.10

Recent research has begun to bring geographic and geometric accessibility measures together. For example, Marcus (2000) proposed the place syntax measures, and Karimi et al. (2013) put forth the origin- and destination-weighted graph-based measures. Place syntax determines access to plots (private) rather than access to streets, which is more akin to geographic measures. Origin-and-destination-weighted centrality measures, on the other hand, borrow the concept of differential origin and destination attraction weights from spatial interaction models but measure accessibility via a spatial dual graph rather than at the plot level. Both of these measures are considered as both geometric and geographic accessibility measures.

3.2.3.2 Distance and Radius

Two key methodological topics that differentiate space syntax research from urban modelling research is the focus on distance and radius. According to space syntax literature, the comparison between different distances is an important focus in spatial cognition research (Montello, 1991; Gibson, 1979; Dera-Abrams, 2006). Space syntax research has shown that a least-angular strategy correlates better with the aggregate pedestrian movement than the metric and the topological distance (Hillier and Iida, 2005; Penn and Turner 2002). This differentiation subsequently allows different questions to be asked; for example, what is the impact of angular distance on house prices?

In space syntax literature, the concept of radius is also an important topic. Radius in space syntax literature is defined as the distance cut-off for an accessibility measure. For example, for each segment, closeness centrality at radius 800 metres measures the segment's inverse average distance to all other segments up to radius 800 metres. Local radius is, therefore, interpreted as neighbourhood accessibility,

and global radius is interpreted as city-wide accessibility. Space syntax research uses the radius parameter to understand cities as dual processes of global-local relations; this creates a different interpretation of cities.

The examination of the global-local relations resulted in the concept of intelligibility, which measures 'the degree to which the number of immediate connections a line has is a reliable guide to the importance of that line in the system as a whole' (Hillier et al. 1987, p.237). Formally, intelligibility calculates the Pearson correlation coefficient, such as connectivity and integration, between a local syntactical property and a global syntactical property. Correlations between spatial relations and socio-economic performance differ between local areas that exhibit varying levels of intelligibility (Hillier et al., 1987).

3.2.4 Discussion and Limitations

To summarise, this section described both geographic and geometric accessibility measures to be used in the first analytical chapter. Geometric accessibility is able to capture spatial network effects that are not considered in geographic accessibility measures. Jointly studying geographic and geometric accessibility effects in the hedonic price model can consequently provide a better understanding of the relationship between accessibility and house prices. There are, nonetheless, some general limitations to accessibility measurement. The first limitation is that accessibility improvements do not necessarily translate into behavioural effects. For example, providing additional public transport infrastructure in a car-dominated city may not necessarily render changes in commuting behaviour. The second limitation is that accessibility measures are inherently static and do not consider capacity and congestion. Therefore, in this regard, it can be argued that transport data or urban simulation models are a more accurate representation than accessibility measures. Despite these limitations, accessibility measures provide a basis for understanding the location potential of a space with limited data. Future research can consider using simulation models or transport data to more efficiently capture the location effect on house prices.

3.3 Research Strand Two: Street-based Local Areas (St-LAs)

The concept of a local area or a neighbourhood is complex and fuzzy, which involves spatial, historical, socio-economic and perceptual characteristics that change and overlap over time and geography (Lebel et al., 2007; Galster 2001; Kearns and Parkinson, 2001). The majority of the existing housing research uses arbitrary census tracts in a hedonic price model to estimate a neighbourhood effect or to identify a housing submarket. This 'arbitrariness' (Goodman, 1985) or 'ad hoc' nature (Orford, 2000) influences the results of the hedonic price model. The second research strand argues that St-LAs correlate with house prices and that St-LAs have stronger effects on house prices than traditional census-based output areas. The reasons are twofold. First, St-LAs are able to more precisely capture the subtle perceptual differences in urban environments than administrative local areas. Local areas have a neighbourhood effect on house prices. Second, over time, the street network topology reinforces socio-economic similarity within a local area. Differences between local areas can become more pronounced,

as like-minded people encounter each other, cluster together and share information with each other. Figure 3.6 illustrates the comparison between an St-LA and a census tract output area for the Isle of Dogs in London.

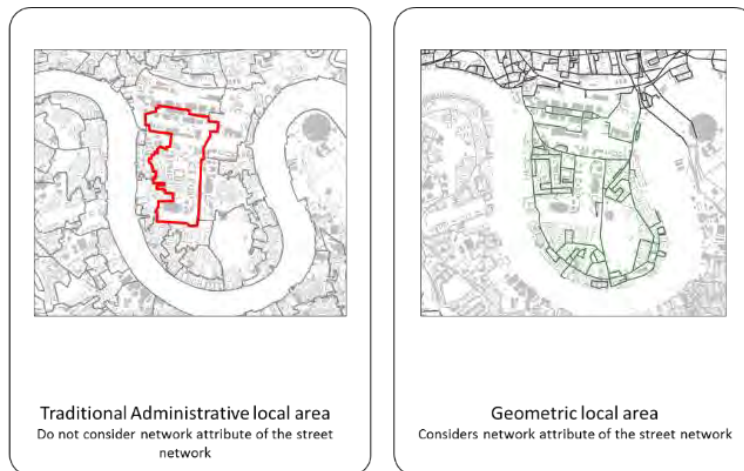


Figure 3.6 On the left is the traditional administrative local area, which does not consider the network attribute of the street network (red) and on the right is the Street-based-Local-Area (green).

3.3.1 Strand Two Method

This section presents the key research method used for studying St-LAs in the second research strand. This research first describes the classification of UK administrative geography, followed by the definition of the St-LA. The UK administrative geography includes the postcode unit, the Lower Super Output Area (LSOA), the Medium Super Output Area (MSOA) and the electoral or statistical ward. The next section describes and defines the St-LA. Various approaches, including traditional syntactical measures and community detection techniques, are compared and discussed.

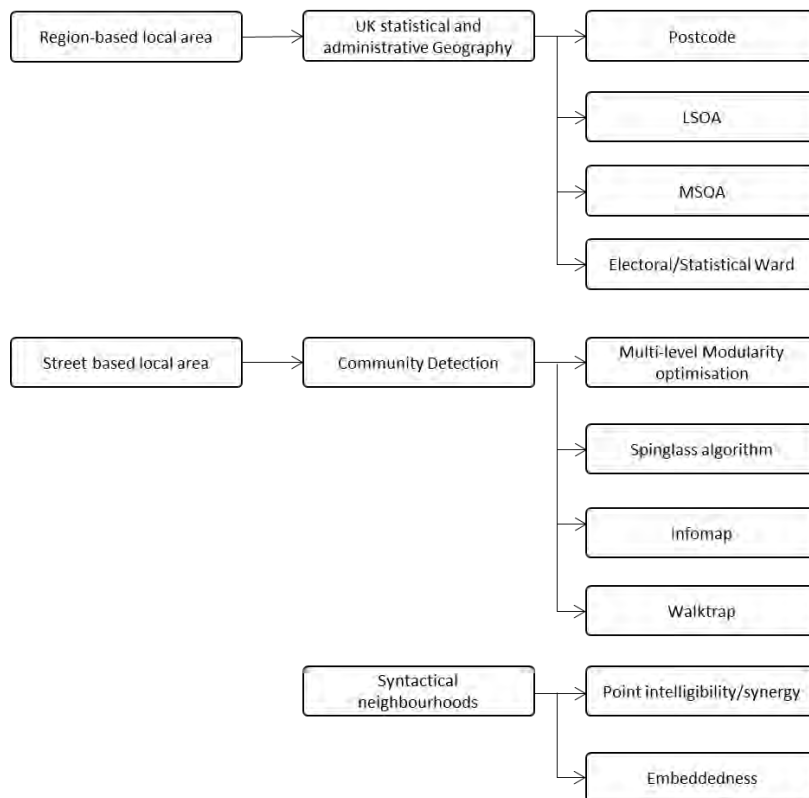


Figure 3.7 The St-LA unit and the traditional region-based local area unit.

3.3.2 UK Statistical and Administrative Geography

There are many different geographical classifications in the UK; the dominant types are administrative, census and postal (ONS 2015). There are other functional classes, such as health and workplace geography. Figure 3.8 conceptualises these three primary classes of geography on the horizontal axis and the size of the units on the vertical axis, where the units at the top are larger and

those at the bottom are smaller.

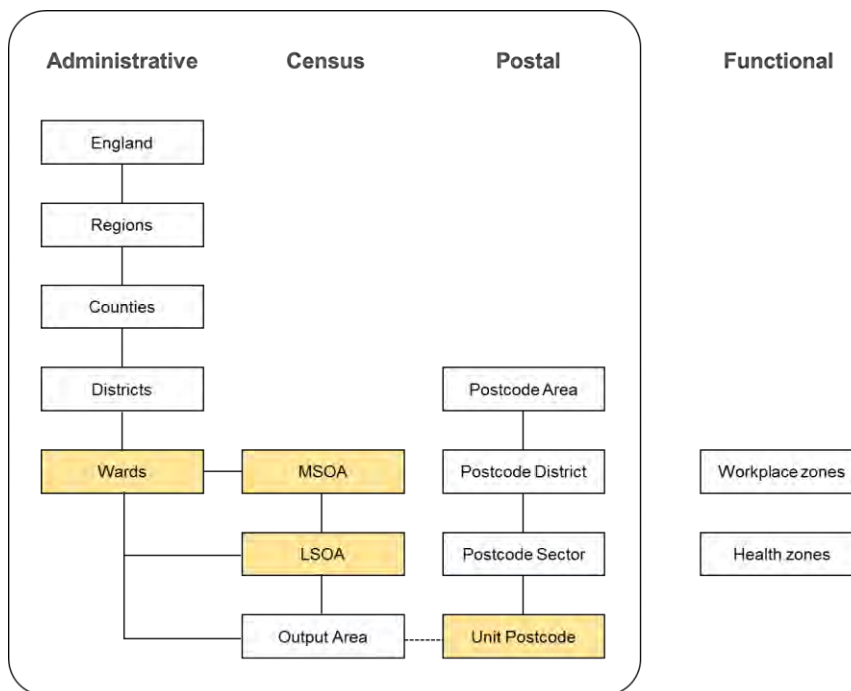


Figure 3.8 The UK administrative, census and postal geography classes (ONS 2015).

The left-hand column shows the administrative classifications of the whole country down to the regions, counties, districts and electoral wards. Administrative geographies are mainly used for governance purposes, such as elections, government functions and policies. Administrative geographies are then connected to the census classifications, which include MSOA, LSOA, and Output Areas (OA). As already explained, these OAs are used for the UK census geography that is undertaken every 10 years. These classifications are then connected to the postcode classification, which is mainly used for postal delivery processes. Postcode classifications include the postcode area, the postcode district, the postcode sector, and the postcode unit, in which these categories are subsequently smaller. Due to the size of the postcode unit, the connections between postal addresses and both the census and administrative geographies are not always seamless. The next section describes four of the geographical definitions highlighted in Figure 10.

- The UK statistical and administrative geography
 - The UK statistical and administrative local area
 - Postcode unit
 - Super Output Area (LSOA/MSOA)
 - Electoral ward

3.3.2.1 UK Statistical and Administrative Local Area

Postcode units, LSOAs, MSOAs and electoral wards were selected for the analytical study. These local areas were selected to represent a geography smaller than a district but larger than the home (Kearns and Parkinson, 2001). Figure 3.9 shows these four types of local areas overlaying the Thamesmead

area of London, from the postcode unit in the top left and the LSOA in the top right, to the MSOA in the bottom left and the ward in the bottom right.

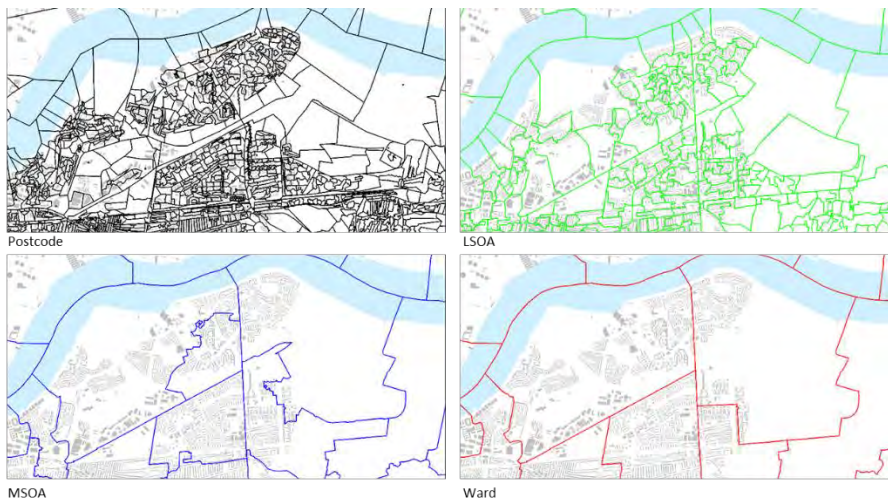


Figure 3.9 Postcode unit, LSOA, MSOA and ward visualised in Thamesmead

3.3.2.2 Postcode Unit

The postcode unit is the smallest geography in the UK after individual addresses. These units are defined by the Royal Mail to identify postal delivery areas in the UK and are generated according to the number of adjacent addresses. There are over 1.75 million postcode units in the UK, with an average of 15 addresses per unit. A postcode unit does not exactly map OAs or wards. The key benefit to using the postcode unit is that it represents the smallest geography in the dataset. However, it is also more arbitrary, as it is defined by the number of addresses it contains. As a result, there are cases where there are multiple postcode units in one building or multiple buildings within one postcode unit.

3.3.2.3 Super Output Area (LSOA/MSOA)

The next level is the super output area (SOA) geographies as defined by the Office of National Statistics (ONS), which are aggregated from the OAs. The OAs are constructed by aggregating postcode units according to multiple criteria, including size, shape, population, tenure and dwelling type. The SOAs are restricted by physical boundaries, such as rivers and roads. According to the ONS (2015), the minimum population for the LSOA tract is 1,000 and 5,000 for the MSOA tract. These minimum sizes are required to ensure the confidentiality of the data. Overall, a total of 32,844 LSOAs in England and 1,909 LSOAs in Wales are generated; correspondingly, there are 6,791 MSOAs in England and 410 MSOAs in Wales. These boundaries change over time, as demographics also change. For example, if the population rises above the threshold for an output area, the output area is split into two. If the population falls below a threshold, multiple OAs are aggregated. A benefit of using SOAs is that the UK census data sits within the SOA geography.

3.3.2.4 Electoral and Census Area Statistic Ward

The final type of administrative geography used in the study is the electoral and census area statistics (CAS) wards. This is the key UK administrative geography used to elect local government councillors.² The population is approximately 5,500 per ward. The CAS wards are a more stable definition of the electoral wards for data collection purposes. There are 8,850 CAS wards in England and Wales, which are used for this study. The benefit of using wards is that it is the smallest and most consistent geography with linkages to census tract output area data.

3.3.3 Street-based Local Areas

This section begins by defining St-LAs. Second, we review previous syntactical neighbourhood measures and their shortcomings when used as St-LAs. Third, we describe which community detection techniques are used to identify the St-LAs. Lastly, we compare different community detection methods and select an appropriate one for the analytical study. The following section is set out as follows:

- Street-based local area
 - Definition
 - Previous syntactic method
 - Embeddedness
 - Point intelligibility or synergy
 - Community detection method
 - Multilevel modularity optimisation
 - Walktrap algorithm
 - Spin glass algorithm
 - Infomap algorithm
 - Comparison between the community detection methods

3.3.3.1 Definition

This section begins by defining an St-LA as a local area that, firstly, is configurational or topological; secondly, is street-based; thirdly, has membership in discrete form; and lastly, is larger than a home area but smaller than a district. The concept of St-LAs borrows from two fields, network science and space syntax research. From network science, St-LAs derive the concept of community structure, where the optimal partition of a connected graph has greater connectivity within the partition than between partitions. These community structures strongly correspond with groupings in different types of social and biological networks (Girvan and Newman 2002). This community detection technique is applied to the spatial dual graph, as borrowed from space syntax literature, to represent the city. This is an innovation for this research and is explained in Section 2.2. The St-LA is defined where each subgraph **sg** is a subset of the edge set **V** in the spatial dual graph **DG=(V,E)**.

² Districts, unitary authorities and London Boroughs in England. Unitary Authorities in Wales. Council areas in Scotland. District council areas in Northern Ireland.

$$\forall i, sg_i \subseteq V, StLA = \{sg_1, \dots, sg_k\}$$

St-LA is street-based local area

sg is the subgraph

Equation 3.11

3.3.3.2 Previous Syntactical Methods

An early enquiry in defining a local area through its spatial morphology emerged from the field of space syntax. One of the earliest observations was made by Hillier et al. (1987). They found that the correlation between spatial configuration and pedestrian movement differ between local areas. Penn (2001) called these local areas ‘correlation detectors’. Peponis (1988) made the observation that highly accessible routes act as natural boundaries between neighbourhoods, and Read (1999) observed that neighbourhoods are often found in places of high local integration. The former observation can be interpreted as neighbourhoods being divided by high movement corridors, while the latter suggests that the heart of the neighbourhood has greater node density than its edges. These emerging ideas led to the conjecture of syntactical local area from Yang’s (2007) embeddedness measures and Dalton’s (2006) point intelligibility measures. The former focused on the node count density differences between two radii. The latter focused on defining a local version of intelligibility to identify syntactic neighbourhoods. Embeddedness **EMD** is defined as the node count difference of two different radii ($r1 \neq r2$) divided by the difference of its radius, as follows:

$$EMD_i(r1, r2) = \frac{N_i(r1) - N_i(r2)}{r1 - r2} \quad \{r1 > r2\}$$

Where

EMD_i is the embeddedness measure of i

r1 is the first radius

r2 is the second radius

N_i(r1) is the node count at i in radius (r1)

N_i(r2) is the node count at i in radius (r2)

Equation 3.12

In addition, point intelligibility *PI* for each node is defined as the Pearson correlation coefficient between global integration and the connectivity at radius *rad*. One can interpret this as a local measure of intelligibility at a certain radius. More formally stated:

$$PI_i(rad) = corr(CC_i(rad), DC_i)$$

Where

PI_i is the measure of point intelligibility at i in radius r

corr is the Pearson correlation between CC and DC

CC_i is the closeness centrality (global integration) at i

DC_i is the degree centrality or connectivity at i

rad is the radius

Equation 3.13

Both research studies found descriptive correspondence to named local areas in London, and it is

argued that spatial configuration not only influences pedestrian movement distribution but also their perception of place. However, this method is descriptive in nature and can only suggest whether an area has lower intelligibility or lower embeddedness. Furthermore, this method cannot show this intelligibility or embeddedness if two spaces are in the same local area. As a result, neither measure is applicable for the analytical study. Future research can transform these descriptive measures into St-LAs by using a cut-off threshold. Instead, this research explores the use of the community detection techniques in defining an St-LA.

3.3.3.4 Community Detection Techniques

This research suggests that community detection techniques can be used on the spatial dual graph in identifying an St-LA. Borrowing from network science, the objective of community detection is to define a set of subgraphs that maximise internal ties and minimise external ties by using the strict topology of the graph. More formally, for all nodes in the community subgraph SG , the number of connections within the cluster K^{in} is greater than the number of connections to the rest of the network K^{out} .

$$K_i^{in} > K_i^{out}$$

K^{in} is the within-cluster degree

K^{out} is the between-cluster degree

Equation 3.14

Due to its usefulness in identifying network groupings, there is a large quantity of literature from a range of disciplines in regards to defining community structures. These can, in general, be organised into five categories (Reichart and Bornholdt, 2007; Raghavan et al., 2007; Newman and Girvan, 2004; Pons and Latapy, 2005). The first category is centrality-based, such as the betweenness cut algorithm, where links with the highest centrality are removed to identify the disconnected component. This algorithm is not applicable for large graphs due to the computation complexity of the betweenness centrality. The second is modularity-based; this includes the modularity optimisation method (Newman 2002), which maximises a quality function, and the Spin glass algorithm, which minimises an energy function. The third is spectral-graph-based, where membership is based on the leading eigenvectors of a graph Laplacian. The fourth is based on a random walk. This includes the Walktrap algorithm, which is centred on the premise that if two nodes are in the same community, the probability of a random walker reaching both is similar. The fifth is information-based. This category includes Infomap (Rosvall and Bergstrom, 2007), in which entropy measures of information minimisation are used to identify an optimal partition.

In the network science literature, performance of community detection algorithms are traditionally assessed either through its partition quality (ie. modularity) or through the speed of the calculation. For example, Lancichinetti and Fortunato (2009) developed partition benchmarks from artificial networks where Infomap, modularity optimisation and Potts spinglass achieved better performance than other methods. Orman (2011) similarly concluded that the best-performing techniques were

Infomap, Walktrap, Spin glass and the modularity method.

These four algorithms, inspired by information theory, graph theory and statistical mechanics, are briefly described in the next section and followed by a descriptive comparison in an urban setting. An extensive description of each method and the motivations is beyond the scope of this research. Rather than testing and comparing all of the community detection techniques, the aim of this section is to select a commonly used community detection techniques from the network science literature appropriate for urban settings.

As there are no commonly used ground-truth labels or benchmarks for such techniques in an urban setting, the selection of the community detection technique is informed by a visual comparison to the named area (“ground-truth”) boundaries used from previous research (Dalton 2007). The selected technique is then used in the analytical chapter. For a detailed description of each of the community detection algorithms, please refer to the corresponding literature.

3.3.3.5 The Modularity Optimisation Algorithm

The first algorithm is the modularity optimisation algorithm, which defines local areas by the within-cluster connectivity of the network (Girvan and Newman 2002). Modularity (Equation 3.15), the most popular quality function used in community detection, calculates the difference between the observed number of edges within a subgraph and the expected number of edges. The greater the observed number of edges relative to the expected number, the higher its modularity. More formally stated: Modularity Q is defined where A is the adjacency matrix, m is the total number of edges in the graph, k_i and k_j are the degrees for vertex i and vertex j . Furthermore, if i and j are in the same community, δ is 1; if they aren't, then δ is 0.

$$Q = \frac{1}{2m} \sum (A - k_i k_j / 2m) \delta(C_i, C_j)$$

Where

Q is modularity index

A is the adjacency matrix

m is the total number of edges

k_i and k_j are the degree for the two subgraphs i, j

δ is a Kronecker Delta function, which equals 1 when its arguments are the same and 0 otherwise

Equation 3.15: Modularity (Q) equation (Girvan and Newman, 2002).

It is currently impossible to use optimisation against the above function to solve for large datasets.³ As a result, a number of approaches have been implemented for finding a near-optimal subgraph (Girvan and Newman, 2002). One method is to apply a hierarchical approach (Clauset et al, 2004; Blondel et al., 2008) to optimise against the modularity function, as illustrated in Figure 3.10.

³ In computation, this is considered a class NP-hard problem.

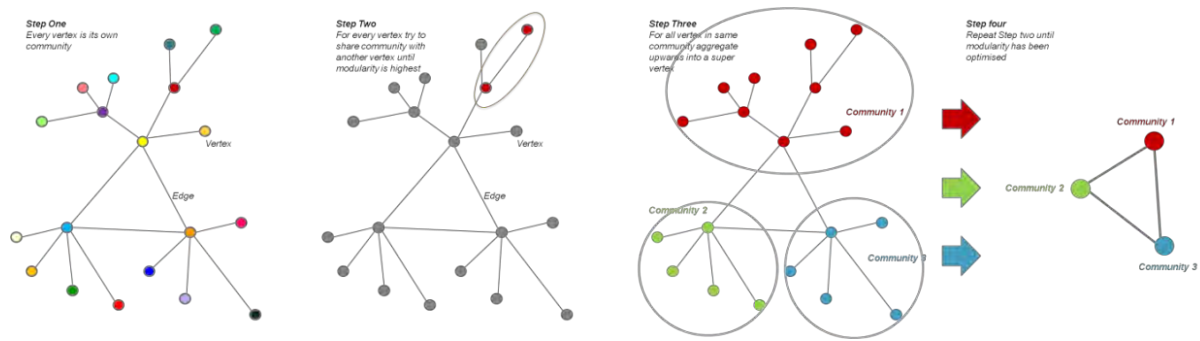


Figure 3.10 The modularity optimisation algorithm. (diagram produced by the author).

The modularity optimisation algorithm starts where every vertex is a subgraph. Every vertex then shares a subgraph membership with the neighbour that attains the highest modularity score. This continues for all vertices. After all of the vertices are traversed, the vertices within the same subgraph are aggregated into a new super vertex. The super vertices again aggregate with their neighbours, and this continues until modularity can no longer be optimised. This approach is hierarchical in nature.

3.3.3.6 The Walktrap Algorithm

The second algorithm is the Walktrap algorithm, which applies the random walk concept to defining local areas (Pons and Latapy 2005). The Walktrap algorithm shows that if two nodes are in the same community, the distance to get to a third node within the same community is also similar.

The first step is to calculate the random walk distances between i and j . If two vertices are in a different community, the distance must be larger, and if they are in the same community, the distance must be smaller. The probability of a random walker at each step moving from i to j is P_{ij}^t .⁴ If two vertices are in the same community, a walker tends to see all of the other vertices in the same way. Pons and Latapy (2005) showed that if i and j are in the same community, we can use its distance r to the third vertex k to calculate its distance:

$$r_{ij} = \sqrt{\sum \frac{(P_{ik}^t - P_{jk}^t)^2}{d}}$$

Where

r is the distance between i and j

P_{ik} is the probability of walking from i to k

P_{jk} is the probability of walking from j to k

d is the degree

Equation 3.16

⁴ $P_{ij}^t = \frac{d_j}{\sum d}$

Where

P is the probability of moving from i to j at t

d is the degree

From these distances, we then choose which communities to merge according to the standard hierarchical clustering analysis (Pons and Latapy 2005). At each step k , we merge two communities that minimise the mean of the squared distance between each vertex and its community. Details of the Walktrap algorithm can be found in Pons and Latapy (2005).

3.3.3.7 Infomap Algorithm

The third algorithm is the Infomap algorithm (Rosvall and Bergstrom 2007). Infomap is also based on the random walk concept; however, instead of calculating distances between vertices, this algorithm detects communities by minimising the information needed to characterise a random walk. This research borrows the concept of Shannon entropy H from information theory. In order to turn the network into information bits, this approach starts by giving unique names to every node and partition in the network. The partition that minimises the information, as denoted by $L(M)$, is then the network's optimal partition. The Infomap equation $L(M)$ is split into two parts. The first part is the entropy $H(Q)$ of the movements between modules Q and the second part is the entropy $H(P)$ of the movement within modules P . In Equation 3.17, $L(M)$ sums up the two parts, which gives the average information (number of bits per step) to describe a random walk partitioned into M communities.

$$L(M) = q_{\sim} H(Q) + \sum p_{\sim}^i H(P^i)$$

Where

$L(M)$ gives the average information of a random walk partitioned into M communities

$H(Q)$ is the entropy between-module movement

$H(P)$ is the entropy within-module movement

q_{\sim} is the probability that the random walk switches module

p_{\sim}^i is the probability that the random walk stays within module i

Equation 3.17

To minimise Equation 3.17, Rosvall and Bergstrom (2007) used simulated annealing to reach the optimal partition. This algorithm has been tested on the citation network, as illustrated in Figure 3.11. For more details, please see Rosvall and Bergstrom (2007).

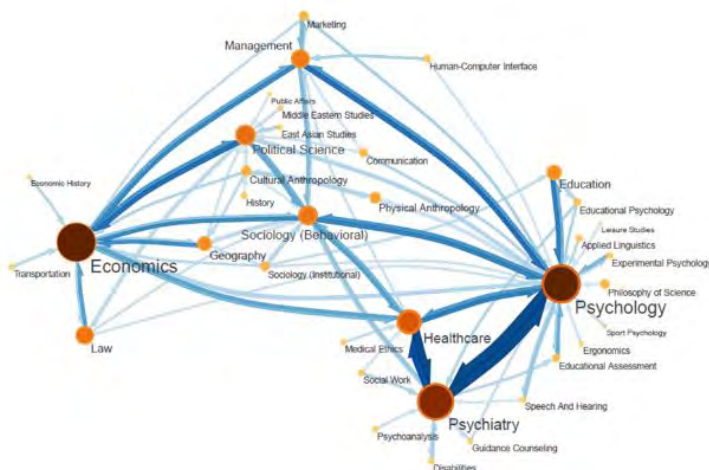


Figure 3.11 A map of the social sciences (Rosvall and Bergstrom 2007).

3.3.3.8 The Spin Glass Algorithm

The fourth algorithm is the Spin glass algorithm (Reichart and Bornholdt 2006). The authors were able to solve the community detection problem by mapping the concept of modularity maximisation as an equivalent minimisation problem in a statistical mechanic model, known as the Potts Spin glass model. The optimal partition of the network is interpreted as the spin configuration that minimises the energy of the system, as denoted below:

$$H(\sigma) = -\sum(A_{ij} - \gamma(p_{ij}))\delta(\sigma_i, \sigma_j)$$

Where

H is the Hamiltonian

A_{ij} is the adjacency matrix

p_{ij} is the probability that a link exists between i and j

γ is a balancing factor

δ is a Kronecker Delta function which equals 1 when its arguments are the same and 0 otherwise.

Equation 3.18

As noted by Reichart and Bornholdt (2006), H is essentially the inverse of the modularity Q in the equation 3.15 when $\gamma = 1$. Thus, finding an optimal partition that maximises the modularity in community detection is the equivalent of finding the ground-state spin configuration that minimises the total energy of the Potts model (Equation 3.18). The author has used a two-part recursive partitioning method to identify the optimal partition, where the total energy H cannot be lowered any more. See Reichart and Bornholdt (2006) for more details about the statistical mechanics concept, proof and implementation.

3.3.3.9 Community Detection Methods Comparison

Previous empirical research focused mainly on the application of community detection techniques to social and biological networks which are non-planar, both spatial and aspatial (Newman 2010). Applying community detection techniques to the street-network dual graph, which is planar and spatial, is therefore a novelty of this research. The primary method used for the test in this section is a descriptive comparison on a set of ground-truth labels of urban neighbourhood boundaries. Community detection techniques is first applied to the street segment segment dual graph of London. Figure 3.12 shows the extent of the study area.



Figure 3.12 The study area boundary.

This section compares both statistically and visually the four algorithms when they were applied to the dual graph of the London's street network segments. Statistically, the Modularity optimisation, Walktrap and Spin glass algorithms produced similarly sized local areas, with more than 500 segments. The standard deviation was also larger in the Walktrap algorithm than in the modularity optimisation and Spin glass algorithms. The Infomap algorithm, however, created clusters made up of only two lines. Parameters of this algorithm might need to be adjusted in order to adapt to spatial network models.

Table 3.1 Statistical comparison between community detection algorithms.

	Walktrap	Modularity	Spin glass	Infomap
Mean	562	549	568	2
Standard Deviation	411	257	134	0
Standard Err Mean	29	18	9	0
Upper 95% Mean	619	584	586	2
Lower 95% Mean	505	513	549	2
N	202	207	200	51,044
Computation Speed	> 3 hours	> 10 mins	> 8 hours	> 1 hour

To compare the algorithms, four named urban areas in London were juxtaposed, including Hampstead Garden Suburb (1906), Bedford Garden Suburb (1875), the Thamesmead development (1960), and the Soho district in Central London (1600).⁵ These areas were selected to represent the distinct named areas in London. This type of visual comparison has been used in previous syntactical studies (Dalton 2006; Yang 2007). These urban developments were tested through four algorithms; namely, the modularity optimisation algorithm, the Walktrap algorithm, the Spin glass algorithm and the Infomap algorithm. For each algorithm, the parameters were selected by visually matching them with the named local area.

⁵ The following sources were used for the identification of the known local area boundary. (LB Barnet, NA; LB Ealing, 2007 & 2008; Thamesmead Trust, 2007; Sheppard, 1966; Walter, 1878; Wikitravel, 2011)

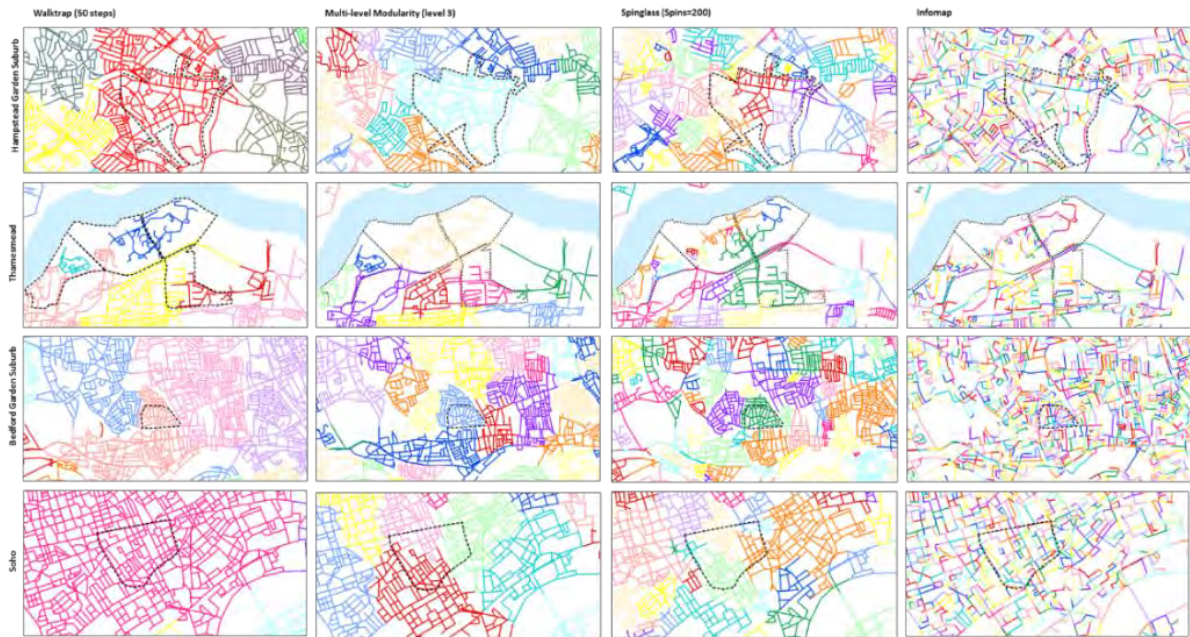


Figure 3.13 A visualisation matrix between the Walktrap, modularity optimisation, Spin glass and Infomap algorithms for the four named urban areas in London.

The visualisation matrix shows the four named areas on the vertical axis and the four algorithms on the horizontal axis. Each colour represents a different membership, with the named local urban area boundary shown in black. The results showed that the Infomap algorithm did not achieve a good fit with any of the urban areas. In addition, the results revealed that the three algorithms were more accurate for the three planned named areas—Hampstead Garden Suburb, Bedford Garden Suburb and Thamesmead—than for the organic neighbourhood represented by Soho. The results were logical, as these three areas were planned by a single developer or architect, whereas Soho has a much more porous and connected grid to the surrounding area, which has developed organically. More specifically, the modularity optimisation algorithm visually had a more accurate partition than the Walktrap and the Spin glass algorithms. This partition was more obvious for Hampstead Garden Suburb and the Thamesmead area. For Soho, the modularity optimisation and Spin glass algorithms vaguely identified St Anne's, the eastern parish in modern Soho, as shown in Figure 3.14. Further statistical tests (F-tests) are calculated in the analytical chapters.



Figure 3.14 St Anne's Church, Soho (London County Council 1966).

The test results showed that community detection methods can be applied to the spatial street network dual graph in order to identify St-LAs, which is a key contribution to the research objective. The modularity optimisation was selected as the appropriate method for the analytical chapter, as this algorithm visually produces more accurate visual partitions with greater computational efficiency than the other methods. The Walktrap algorithm, however, is a close alternative; the Infomap and Spin glass algorithms had less accurate results. Due to the small sample size of this pilot study, further research is required to adapt these algorithms to an urban setting. To conclude, the results from the community detection techniques were visually more accurate in identifying planned local areas than organic ones. This result suggests that the formation of the local area is a complex subject, in which spatial configuration provides only one perspective.

3.3.4 Discussion

To summarise, this section applied modularity optimisation to the spatial network dual graph to define the St-LAs used in the second analytical chapter. The key reason for the adoption is that the modularity optimisation technique can identify the St-LAs on a spatial dual graph with significant visual accuracy while maintaining high efficiency. There are a number of benefits to utilising the street-based perspective in defining local areas. First, using the street network as the geographic unit reduces the modifiable areal unit problem. Second, the street network provides the possibility of studying spatial and perceptual qualities that were not previously available. Third, the street network is clearly the most permanent of all morphological elements; its slowness allows data to be consistently compared across time, while at the same time, the street network is dynamic enough to reflect the changes in its morphology. However, the definition of an St-LA is not without its concerns. The first is that considering only street network connectivity provides an entirely one-dimensional approach to defining a local area. When a grid is highly uniform and connected, street network connectivity may not be adequate in defining a local area. Consequently, further research is needed to test the extent to which other constructs, such as morphological, sociological, economical and historical characteristics, influence local area formation. Second, more research is also required to examine how St-LAs can improve the definition of housing submarkets; this is discussed in the next research strand. Third, this section used a basic method to identify an algorithm for defining St-LAs in a named area with quality visual accuracy. Future research is needed to address this limitation by making comparisons across a greater number

of algorithms, quantitative methods and case studies. Despite these limitations, the definition of an St-LA provides a novel contribution to using community detection techniques on a spatial dual graph.

3.4 Research Strand Three: Street-based Housing Submarkets (St-HS)

Grisby et al. (1987) defined housing submarkets as units that are reasonable substitutes for one another but relatively poor substitutes for units in other submarkets. Despite a general consensus on their existence and on the statistical tests to infer submarkets, there is a general disagreement on their identification method or the variables to include (Watkins, 2001). The vast majority of existing housing submarket research relies on ad hoc region-based administrative local areas, such as census tracts, to build housing submarkets. These unclear definitions have led to inconsistent results. The third research strand, therefore, argues that housing submarkets defined by combining structural, location, and amenity attributes are associated with house prices and are preferable to those defined by census tracts. This comparison is illustrated in Figure 3.15, in which an St-HS on the left is compared with a traditional housing submarket on the right, which does not consider street network effects.

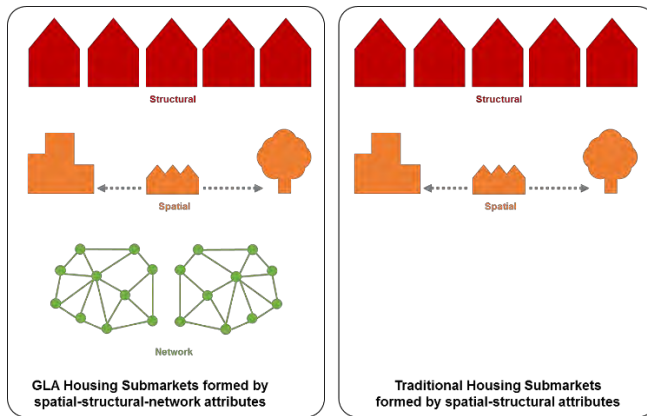


Figure 3.15 On the left is the St-HS that combines street network attributes, spatial attributes and structural attributes; on the right are traditional housing submarkets formed by spatial-structural attributes.

3.4.1 Strand Three Method

This section presents the key research methods used for the third research strand to identify housing submarkets. The submarket construction is based on a three-step process. First, select a geography or local area; second, calculate the averages of the structural, location and amenity characteristics within each local area; and third, employ statistical clustering to each local area to identify the housing submarkets. This research adopts the standard k-means clustering algorithm to define the housing submarkets. These steps are illustrated in Figure 3.16.

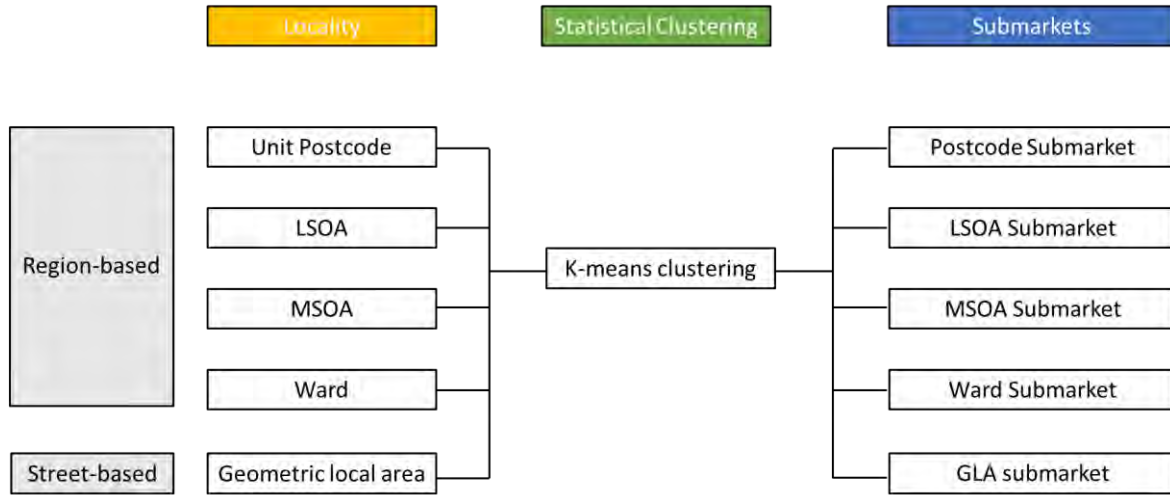


Figure 3.16 The submarket construction process.

3.4.2 The Housing Submarket Definition

The housing submarket is generally defined as a subset of properties whose implicit prices are statistically similar to each other (Schnare and Struyk, 1976). These can be determined a priori (Bourassa, 2002) or empirically driven (Strasheim, 1975; Allen et al., 1995). Standard statistical tests are normally used to ensure that implicit prices are equal within the same housing submarket. This test is known as the homogeneous attribute price vector condition **HAPV**. Due to this condition, housing submarkets are usually larger than neighbourhoods. In this research, each housing submarket **hs** is a subset of all the properties in the housing market **H**;

$$\forall i \text{ } hs_i \subseteq H, HS = \{hs_1, \dots, hs_k\}$$

H is the housing market

hs is the housing submarket

Equation 3.19

where the following homogeneous house price condition is satisfied for all properties **i,j** in the same housing submarket:

$$HAPV: P(z_i) \simeq P(z_j)$$

Where

HAPV is the homogeneous attribute price vector condition

P is the house price

z is a property attribute

i,j are properties where $i \neq j$

Equation 3.20

3.4.3 Statistical Clustering and Algorithms

Due to improved computation abilities and data availability, the statistical clustering method appears to

be the most commonly used method to identify housing submarket clusters in local areas. For example, Goodman and Thibodeau (1998) used hierarchical clustering to define five housing submarkets in Dallas. The Greater London Authority (2004) employed k-means clustering techniques on socio-economic and housing characteristics to define five to six distinct housing submarkets for London's metropolitan region. Kauko et al. (2002) utilised artificial neural networks to identify housing submarkets in Amsterdam.

This study adopts one of the most commonly used classification techniques, known as the unsupervised k-means clustering method, for submarket cluster identification (MacQueen, 1967; Bourassa, 1999). This technique is used to define housing submarkets in the analytical chapter (MacQueen, 1967; Bourassa, 1999). The employment of k-means clustering allows the study to focus on comparing the results across local area units rather than comparing different clustering methods. Thus, the comparison of clustering methods is not the aim of this research.

3.4.3.1 The K-means Clustering Algorithm

The k-means clustering algorithm aims to partition n observations into k clusters that minimise the differences between property attributes. K-means clustering (Equation 3.21) is calculated as follows. Given a set of observations (x_1, \dots, x_n) where each observation is a y -dimension vector, k-means clustering partitioned n observations into $k = \{1, \dots, n\}$ sets of S that minimises the within-cluster sum of squares. The standard k-means clustering algorithm (Lloyd 1965) uses an iterative method, which starts by adopting an initial set of k-means. The algorithm then optimises according to the within-cluster sum of squares until it converges, in which a new assignment no longer changes the sum of squares.

$$\operatorname{argmin}_s \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2$$

Where

x is values for the factor

μ is mean for the factor

Equation 3.21

3.4.4 Limitations

This research strand adopts the unsupervised k-means clustering algorithm for submarket identification. There are some benefits to this method. Being one of the most commonly used clustering algorithms in housing submarket literature, k-means clustering offers a robust and proven method for identifying submarket clusters from multi-dimensional data. However, there are many well-known limitations of the clustering algorithm. As the aim of the research is using a standard method in comparing different local area units, using the k-means clustering algorithm is justified. In the future, other methods in aggregating housing submarkets should be considered, including the use of the fuzzy logic method (Helbich, 2013; 2015) and the machine learning method (Kauko, 2002) for a potentially more accurate identification of housing submarkets.

3.5 Summary

To conclude, this chapter introduced various spatial and statistical methods for the three research topics. This included the geographic and geometric accessibility measures for the first research strand, community detection methods to identify the St-LA for the second research strand and the use of the K-means clustering method to identify the St-HS for the third research strand. The next chapter presents the research framework, the datasets and the case study for the three analytical chapters.

Chapter 4

The Research Framework, the Case Study and the Datasets

4.1 Background

The previous chapter described various spatial configuration methods to measure urban form, which will be used in the analytical chapter. This chapter introduces the research framework, the case study and the datasets used to answer the research question and is divided into three parts. The first segment describes the hedonic price regression model and the multi-level framework, which form the basis of the empirical approach taken in the research project. The second section discusses the research case study. Finally, the third portion of the paper introduces the datasets used for the research.

4.2 The Hedonic Price Regression Model

In order to answer the research question, this study proposes utilising the hedonic price approach as the overarching empirical strategy for isolating the spatial network configuration effects on house prices. As mentioned in the previous chapters, the hedonic price approach uses a regression model to estimate the implicit price of a housing characteristic from the observed sale price (Rosen, 1974). The model in Equation 4.1 shows that the house price is a function of the house's utility-bearing structural, location and neighbourhood characteristics (Freeman 1979). The next section illustrates the multi-level analytical framework used in the thesis. Please refer to Chapter 2 for a more detailed discussion on the theoretical framework and the limitations of the hedonic price model.

$$HP_i = F(S_i, L_i, N_i)$$

where

HP_i is the house price

S_i represents a vector of the structural variable

L_i represents a vector of the location variable

N_i represents a vector of the neighbourhood variable

Equation 4.1

4.2.1 Multi-level Analytical Framework

Previous research suggests that the housing market is hierarchical in nature (Jones and Bullen, 1993; Orford, 1999; Orford, 2001). Orford's empirical research found that the housing market consists of three levels, namely the property level, the neighbourhood level and the district level (Orford, 1999; 2001). This research adopts a similar three-level empirical framework to examine the relationship between the spatial configuration and house prices. Figure 4.1 conceptualises this hierarchical framework of the housing market, where house price variations are represented at three levels: the property level, the local area level and the housing submarket level.

$$HP_{ijk} = F(Prop_i, Local_j, Sub_k)$$

Where

HP is the house price

Prop_i is the property predictor effect which consists of (Li, Si, Ni)

Local_j is the local area effect

Sub_k is the submarket effect

Equation 4.2

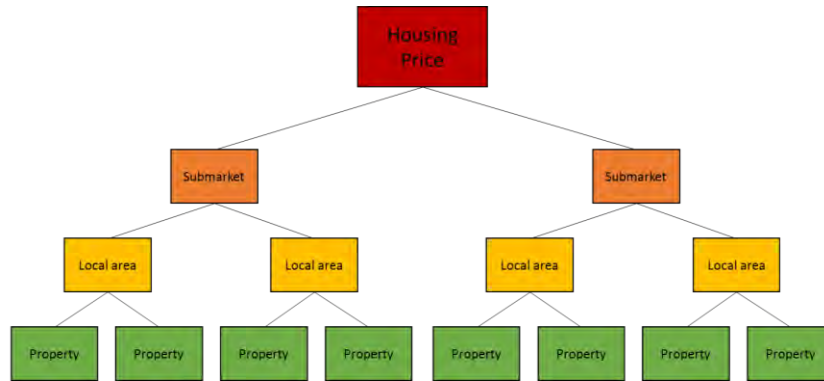


Figure 4.1 The hierarchical nature of the housing market.

These three levels are divided into three analytical chapters. Chapter 5 analyses the geometric accessibility effect at the property level. Chapter 6 examines the street-based local area (St-LA) effect at the local area level. Chapter 7 investigates the street-based housing submarket (St-HS) effect at the submarket level. The detail specifications for the hedonic price regression models are described in each analytical chapter.

4.2.2 Software

This research uses various software, information systems and programming libraries. The spatial network analyses are calculated in QGIS, depthmapX (Varoudis, et al., 2012) and Python (Rossum, 2007). The statistics and regression models are calculated in STATA, JMP, Python and R. By using the GitHub repository, this research plans to open source the scripts used for the research.

4.3 Case Study

Greater London, UK, has been selected for the case study in the analytical chapters. In the following segments, the city of London and its location, government, demographics, economy, transport, and, most importantly, its housing market will be discussed. Section 3 ends with a discussion concerning the strengths and the weaknesses of using the London housing market as the research case study.

4.3.1 Location

For the last two millennia, London has been the political, cultural and economic capital of the UK. Situated along the River Thames, the conurbation of inner London, which grew from the City of London and the City of Westminster, is now a 21st-century global city. Figure 4.2 shows the boundary

of Greater London, separated into inner and outer London, which covers a total of 1,572 square kilometres.

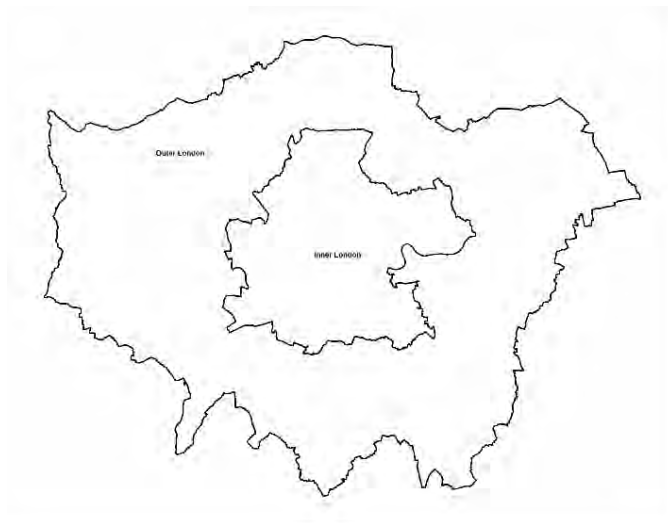


Figure 4.2 The Greater London boundary.

4.3.2 London's Government

The London administration is comprised of two levels, which are a city-wide government body known as the Greater London Authority (GLA) and the 33 local borough government bodies. The GLA can be further divided into two components, which are the Mayor of London and the 25-member London Assembly. The GLA and the 33 local boroughs are jointly responsible for coordinating the city's strategic planning, housing, transport, crime, and fire and emergency planning units. The local authorities consist of 32 local boroughs and the City of London Corporation; they are responsible for local services such as planning, education, social services and infrastructure. Figure 4.3 shows the boundary of the 32 local boroughs and the City of London. The GLA boundary is used as the key study area in the analytical chapter.



Figure 4.3 The 32 London boroughs and the City of London.

4.3.3 London's Demographics

London grew from a population of over one million inhabitants in the early 19th century to a metropolis of over 8.5 million residents in 2011. London's functional definition stretches far beyond the Greater London boundary and, according to the 2011 Census conducted by the Office for National Statistics (ONS), the Greater London built-up area has a population of close to 10 million inhabitants. Figure 4.4 exhibits London's population from 1801 to the present day, illustrating the growth of Outer London since 1891 and the drop in the population in the 1960s, 1970s and 1980s due to post-war construction, the clearing of housing and deindustrialisation. The population is currently increasing, reaching the peak that was attained prior to the Second World War.

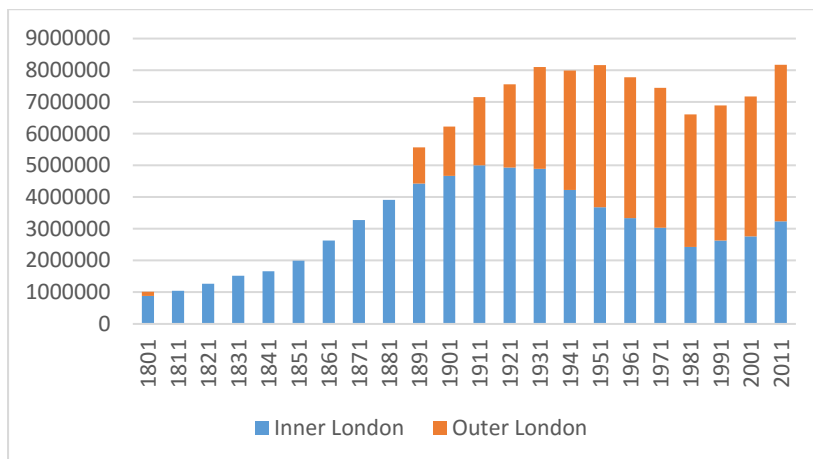


Figure 4.4 London population between 1801 and 2011.

London is also one of the most ethnically diverse cities in the world. According to the ONS, in 2011, 60% of the inhabitants were of white descent, 21% of Asian descent, 16% of black descent, and the remainder being of mixed race descent⁶.

4.3.4 London's Economy

London is the primary economic centre in the UK, with an estimated Gross Value Added (GVA)⁷ of 310 billion GBP in 2012. Economically, the city grew from an industrial powerhouse in the 19th century into a global financial centre in the 21st century. This shift from the industrial sector cannot only be attributed to cheaper labour costs elsewhere, but also to the focus on the service sector-led economy due to advances in information technology, transport and communication (Castells, 2010). Using data from the 2011 UK Census, Figure 4.5 shows the various employment sectors. The figure demonstrates that the service industry, coloured blue, provides over 85% of the total employment in Greater London. The financial sector is London's largest industry, employing over 300,000 (25%) residents. The three non-service industries, namely construction, manufacturing, agriculture and energy, are represented by the colour yellow.

⁶ Mixed is an ethnicity category that has been used by the UK's Office for National Statistics since 1991.

⁷ An economic measure of productivity from individual sector (ONS 2011).

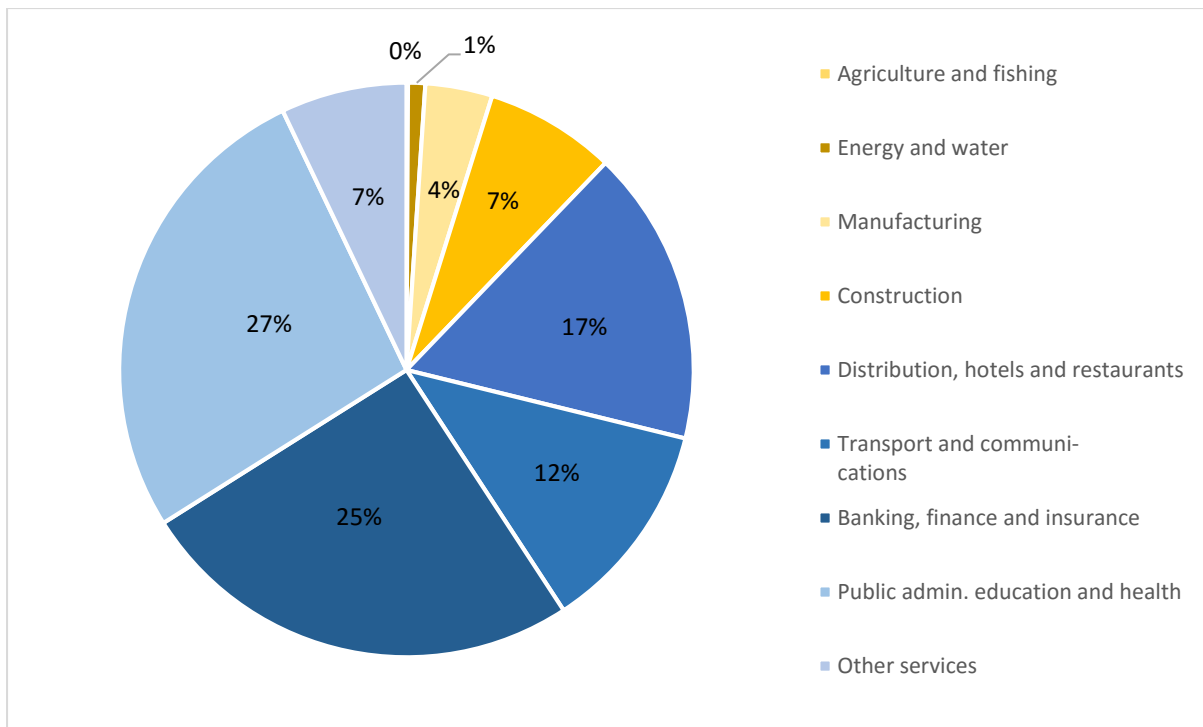


Figure 4.5 London industry sectors in 2011.

4.3.5 London's Transport

"London became a greater and still greater accumulation of towns."

Rasmussen, 1934

The quote above best illustrates the dynamic process of how London became a metropolis, where villages that vary in scale, size and character agglomerated over time to form an interdependent network (Hillier, 1999; 2006). The reduction in travel time due to faster modes of transport allowed this clustering to happen. Examples of these enhanced transport methods include the North and South Circular (London Inner Ring Road) and the M25 motorway (London Orbital), which were built after the Second World War and vastly improved the capacity of individuals able to move across the city, and the world's first railway and underground subway system, which connected numerous cities and villages. These new forms of public transport have played an important role in the city's spatial economy. Figure 4.6 shows the street network in grey, the National Rail system in blue and the London Underground scheme in red; these linkages are overlaid on a map of Greater London's boundaries. The figure exhibits a greater density of public transport in areas of higher centrality and lower density in outlying regions. The direction of dependence between public transport accessibility and urban density is complex (the chicken-and-egg problem) and is subject to further discussion (Davidson, 2008).

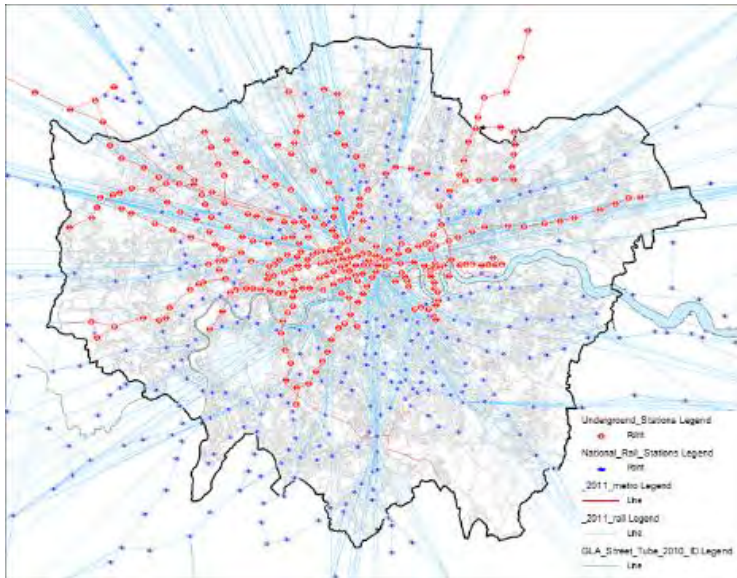


Figure 4.6 The London Underground, the National Railway system and the street network. Author produced the diagram.

Figure 4.7a shows the modal split in the city between 2005 and 2013. The city's modal share can largely be grouped into three transport categories: public transport, private transport and walking or cycling. Figure 4.7b displays the percentage change in each of the modes. A general increase can be seen in five categories: walking or cycling, the London Underground, the National Rail, buses and taxis. During the same period, a general reduction can be observed in two categories, namely private car drivers and motorcyclists. Although not shown here, changes in the inner city are greater when compared to the outer city. In the future, the demand for sustainable modes of transport, such as walking, cycling and public transport, will likely continue to grow. This growing demand is reflected in transport policy projects, such as the construction of the Crossrail project, the expansion of the London Overground, the extension of the Bakerloo Line and the increased provision of cycling and pedestrian infrastructure.

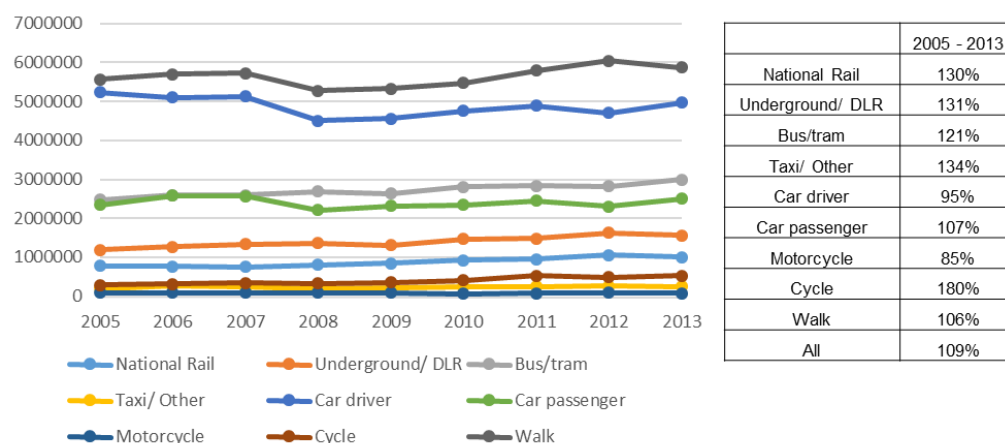


Figure 4.7 London transport modes

- Share of London transport modes between 2005 and 2013 (TfL, 2015) on the left
 - London transport mode share differences between these time periods (TfL, 2015) on the right.
- Source: Transport for London

4.3.6 Discussion

To summarise, London is the political and financial capital of the UK, with over 8 million inhabitants spread across both sides of the Thames. London's growth was made possible due to the city's strategic location, government structure, economic profile, extensive transport infrastructure and, most importantly, human capital. The next section describes the London housing market, which is the focus of the research.

4.4 The London Housing Market

Housing represents one of the largest areas of household expenditure and land use in the UK. In 2011, London had a total of 3.3 million properties, accommodating 8.2 million inhabitants and approximately 2.5 people per household. This average ranges from the City of London's 1.6 people per household to 3.1 people per household in the Borough of Newham (Census, 2011). Figure 4.8 shows that 53% of properties in London are flats, 27% are terraces, 13% are semi-detached and 4% are detached. The average floor area of a London property is 80 square metres; 30% of the London properties have two bedrooms, 28% have three bedrooms and 22% have one bedroom. Most of the properties in London were built before the Second World War (56%), as shown in Figure 4.9 and Figure 4.10. The following section will describe the London housing market concerning three topics: house prices, the London housing crisis and the future of London's housing.

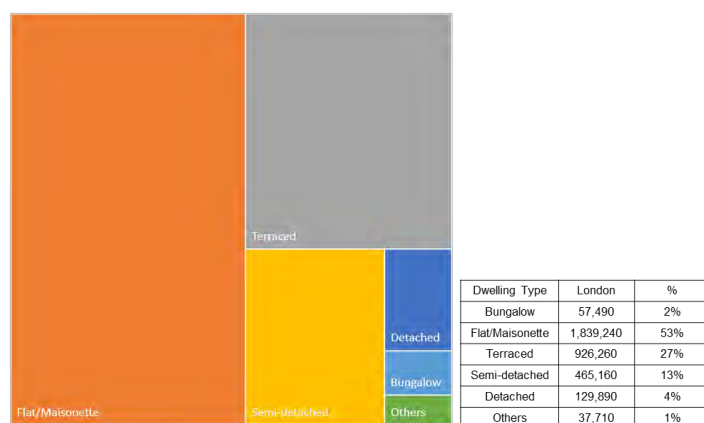


Figure 4.8 Dwelling type in Greater London.

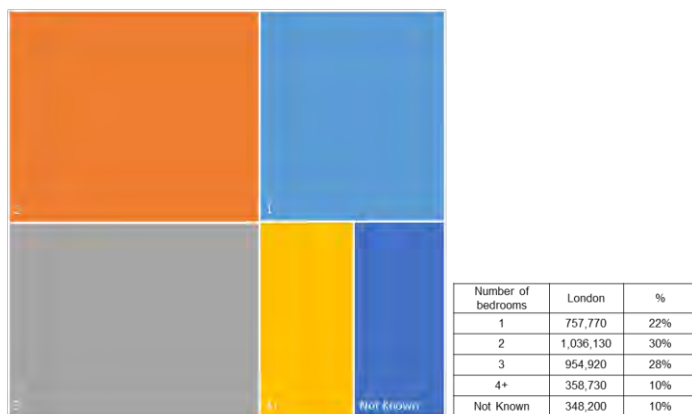


Figure 4.9 Number of bedrooms in Greater London.

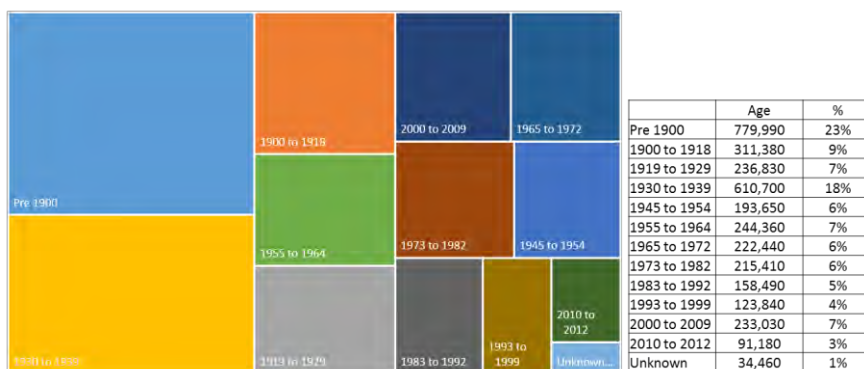


Figure 4.10 Dwelling age in Greater London.

4.4.1 London House Prices

In the UK, house prices have risen significantly over the past 15 years, disproportionately clustering around dense urban areas, such as London. This price increase is primarily due to both the lack of housing supply and the growth in housing demand (Hilber and Vermeulen, 2010; GLA, 2015a; 2015b; Edwards, 2015). Figure 4.11a shows the house price increases in London (blue) relative to the national average (orange), and Figure 4.11b demonstrates the changes in the sales volume in London (blue) relative to the national average (orange). Observations include the divergence in London's house prices with the rest of the country, and the remarkable drop in the sales volume in Greater London since the 2008 financial crisis.

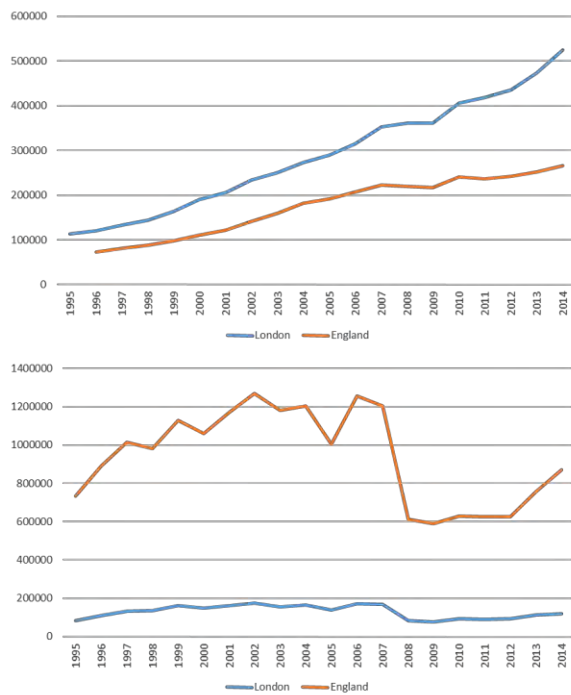


Figure 4.11 London house price and transaction data

a. London house price data

b. London house transaction volume

4.4.2 The London Housing Crisis

The dramatic rise in London's house prices has since led to a housing crisis (GLA, 2014; 2015a; 2015b). Figure 4.12 shows the number of news articles from 2010 to 2015 where 'London Housing Crisis' appeared in the title, as recorded by Google News (2016). The line chart illustrates a dramatic increase in the utilisation of this search phrase since 2013. Prior to 2010-2015, the term 'housing crisis' was related to the sub-prime mortgages and the housing bubble in 2008.

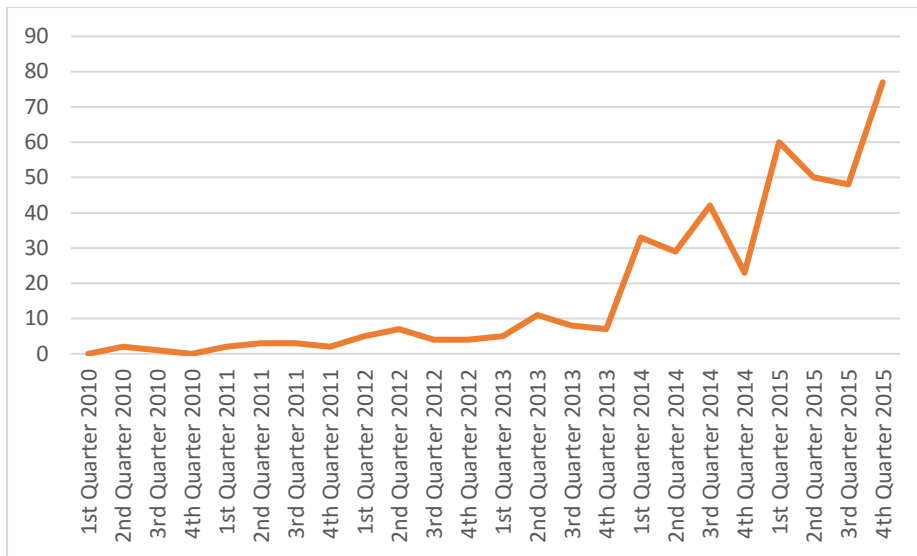


Figure 4.12 London Housing Crisis' appearance on Google search forms from 2010 to 2015. Source: Google, 2015

This rise can be partly attributed to an increase in the demand due to population growth, demographic changes, lifestyle preferences for living or working in the centre, the gravitational shift of London's employment towards the centre and peak car use (Ehrenhalt, 2012; Florida, 1995; Newman and Kentworthy, 2011). Figure 4.13 demonstrates the population changes from 1801 to 2011. The result shows a fall in the population between the 1950s and the 1980s; the population has been growing ever since. London's population has only recently overtaken the peak that was reached in 1939.

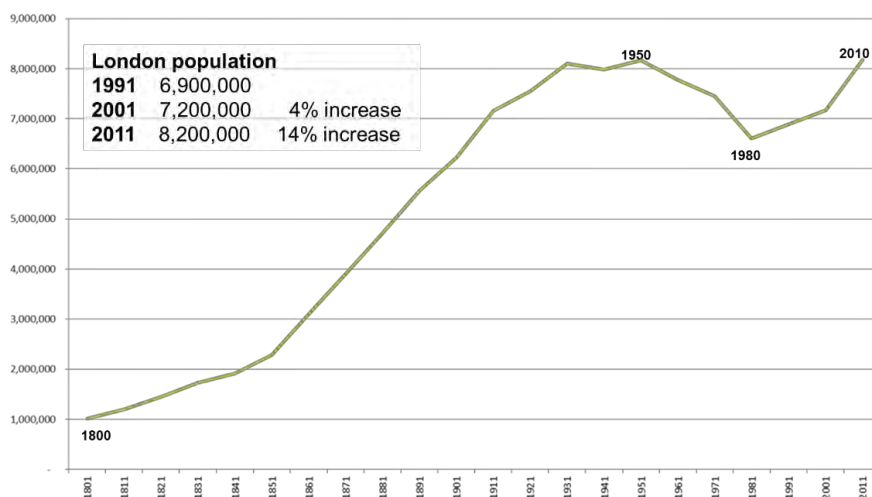


Figure 4.13 Greater London's population between 1801 and 2011.

Figure 4.14 maps the 2011 population and employment distribution and, directly below are the maps depicting the changes from 2001 to 2011 for the same datasets. One observation is the increase in jobs in Central London and the population growth in East London between 2001 and 2011.

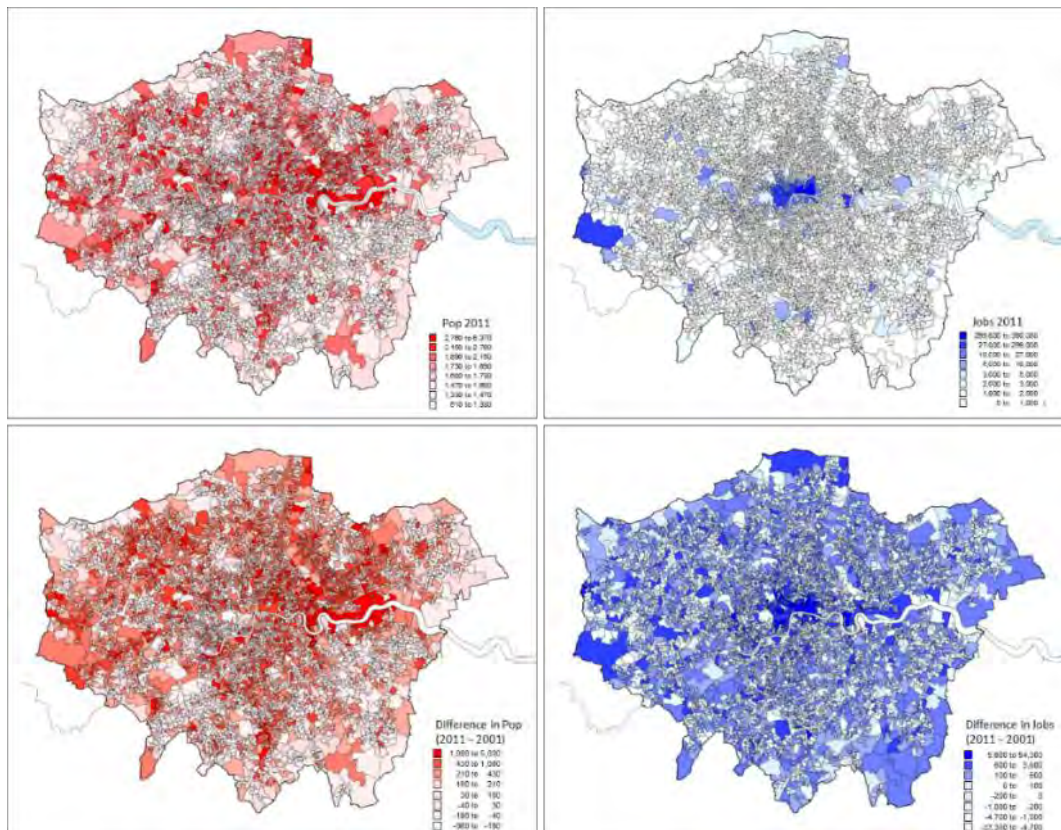


Figure 4.14 Greater London Population and Employment Distribution.

- Greater London population in 2011
- Greater London employment in 2011
- Greater London population growth between 2001 and 2011
- Greater London employment growth between 2001 and 2011

Author produced the diagrams.

The rise in house prices can also be attributed to a lack of supply (GLA, 2014; 2015a; 2015b; LCCI, 2014), where the growth in jobs and population between 2009 and 2014 have not been matched by the number of homes created. Figure 4.15 shows the percentage change in employment and population relative to the housing being built.

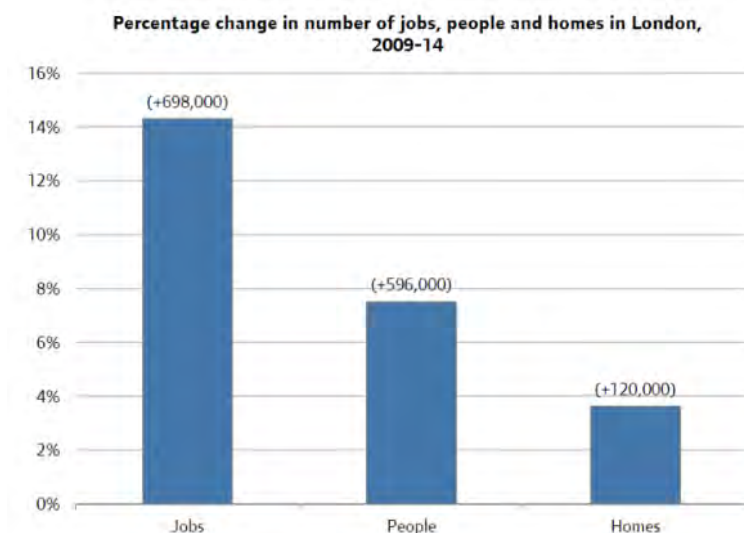


Figure 4.15 The percentage change in the number of jobs, population and homes from 2009 to 2014.

The low interest rates of 0.5% (BOE, 2016) and the stable growth in house prices attracted both local and foreign investors. The compound effect of these factors contributed to the housing crisis observed in London today, where lower income populations are facing rising rents and a lack of affordable housing. Inequality currently presents one of the most prominent challenges for the city (Dorling and Pritchard, 2010). Figure 4.16 shows that in the past 20 years, the ratio between London's median house prices and income, a standard measure of affordability, has increased almost threefold.

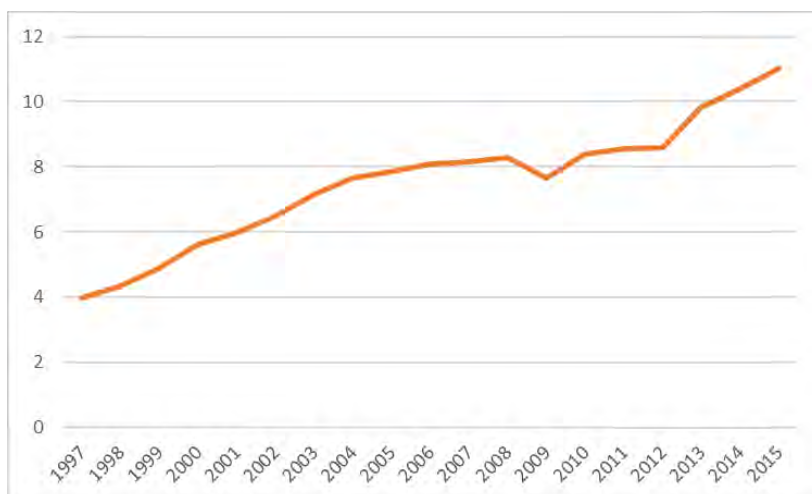


Figure 4.16 The median house price to earnings ratio in Greater London. Source: Department for Communities and Local Government

4.4.3 The Future of London's Housing

Mechanisms and changes in housing policies have been proposed for the next 20 years to support homeownership, increase housing supply, improve existing supply and reduce house prices (GLA, 2014; 2015a; 2015b). According to the London Plan (GLA 2014),⁸ the London Plan alteration report (GLA, 2015b) and the London Housing Strategy (GLA, 2014), 40,000-60,000 homes will need to be built every year to accommodate the 10.11 million inhabitants who will be living in the Greater London area by 2036. The GLA stipulates that new housing in London is focused on five broad categories, including brownfield sites, town centres, opportunity areas (Figure 4.17) and growth corridors, mixed-use developments and the renewal of existing residential areas. Of the total provision, 30-40% is envisaged as being affordable homes, either as low-cost homes or low-cost rents (GLA, 2014; 2015b). Despite these policies to alleviate the negative impacts of the housing crisis, one limitation (GLA, 2014; 2015a; 2015b) is the lack of evidence on how this housing can be delivered and the urban design that is required in building these homes. This gap brings with it a need to better understand how urban form, and particularly its spatial configuration, influences residential location preferences and house prices.

Map 2.4 Opportunity and Intensification Areas

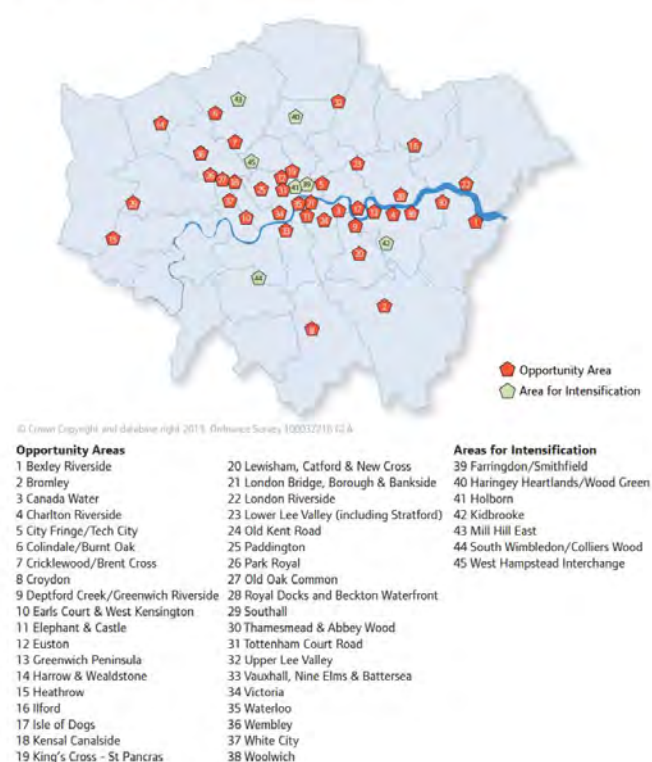


Figure 4.17 London opportunity areas and areas for intensification. Source: GLA (2015).

⁸ London's key spatial development strategy, setting out the social, economic, and environmental framework for the city over the next 25 – 35 years.

4.4.4 Discussion

This research has selected the Greater London housing market for the case study for various reasons. The first is the urgent need to build nearly 40,000-60,000 homes per year to alleviate the housing crisis in London. According to the London Chamber of Commerce, this crisis affects not only the affordability but also the competitiveness of the city (LCCI, 2014). Key components to solving these issues are to build in designated opportunity areas and to create new transport links. These solutions require more evidence on how the street and infrastructure networks influence house prices. The second reason for the selection of Greater London is the growing spatial inequality in London, where evidence is required to better understand these processes in space. Lastly, Greater London has been selected due to data availability.

There are limitations to using London as the research case study. Firstly, London's housing market is complex due to the fact that stakeholders consist of not only homeowners and tenants but also local investors and global institutions and companies. To disentangle this complexity is beyond the scope of this research. Instead, this research focuses solely on providing evidence of how spatial configuration influences intra-city house price variations. The second limitation is the short timeframe of the study. Over the past 15 years, London's housing market has, in general, been on the rise. As a result, this research is representative of only this time frame. The stability of the results can be challenged where the spatial differentiation of house prices is consistently increasing due to the growth in city size and trip length (Banister, 2007; Edwards, 2015). The third limitation is the selection of only one case study across time. Further evidence is required to generalise the findings across a longer period and over different geographical regions. The next section describes the datasets used for the analytical chapters.

4.5 Variables and Dataset

Table 4.1 illustrates the datasets and the variables used for the three analytical chapters. The table includes the house price dataset as the dependent variable and the several datasets utilised to calculate the various structural, location, local area and submarket independent variables.

Table 4.1 Hedonic price model dataset and variables specification.

Category	Type	Variable	Dataset	Source	Research Strand		
					1	2	3
Dependent Variable	Property	Sold Price	Property Dataset	Land Registry, Nationwide			
		x,y	Property Dataset	Land Registry, Nationwide			
		Date	Property Dataset	Land Registry, Nationwide			
Independent Variable	Location Variable	Gravitational Potential to Employment	ONS LSOA Dataset	Office National Statistics			
			Meridian Line Network Dataset	Ordnance Survey			

			TfL Rail Network Dataset	Author			
		Harmonic Centrality	Meridian Line Network Dataset	Ordnance Survey			
			TfL Rail Network Dataset	Author			
		Closeness Centrality	Meridian Line Network Dataset	Ordnance Survey			
			TfL Rail Network Dataset	Author			
		Betweenness Centrality	Meridian Line Network Dataset	Ordnance Survey			
			TfL Rail Network Dataset	Author			
	Local Area Variable	Postcode Unit	ONS Postcode Unit Dataset	Office National Statistics			
		Ward	ONS Ward Dataset	Office National Statistics			
		LSOA	ONS LSOA Dataset	Office National Statistics			
		MSOA	ONS MSOA Dataset	Office National Statistics			
		Geometric Local Area	Meridian Line Network Dataset	Ordnance Survey			
	Submarket Variable	Postcode Unit Submarket	Various Datasets*	Various*			
		Ward Submarket	Various Datasets*	Various*			
		LSOA Submarket	Various Datasets*	Various*			
		MSOA Submarket	Various Datasets*	Various*			
		Street-based Submarket	Various Datasets*	Various*			
	Structural Variable	Dwelling Type	Property Dataset	Land Registry, Nationwide			
		Dwelling Size	Property Dataset	Land Registry, Nationwide			
		Dwelling Age	Property Dataset	Land Registry, Nationwide			
		Tenure	Property Dataset	Land Registry, Nationwide			
		Number of Bedrooms	Property Dataset	Land Registry, Nationwide			
		New Built	Property Dataset	Land Registry, Nationwide			
	Neighbourhood Amenity	Retail Attraction within 800m	Retail Address Point Dataset	VOA			
		Distance to Parks	Park Dataset	Park			
		Secondary School Performance	School Performance Dataset	Department of Education			

The next section describes the key datasets for the research. As most of the datasets appear in more than one chapter, these datasets are described in the order of the analytical chapters and do not appear more than once.

- Research Strand One dataset
 - House price dataset
 - Street and TfL network dataset
 - Retail address dataset
 - UK school performance dataset
 - London heritage parks and gardens dataset
 - Census employment dataset
- Research Strand Two dataset

- Ward dataset
- Super output area dataset
- Postcode unit dataset
- Street network dataset
- Research Strand Three dataset

4.5.1 Research Strand One Dataset

Six datasets are described in the first research strand, namely the house prices dataset, the street and TfL Tube network dataset, the London retail address dataset, the UK school performance dataset, the London heritage parks and gardens dataset and the Census employment dataset.

4.5.1.1 The House Price Dataset

Two London house price datasets are used as the key dependent variables for the three analytical chapters of the thesis. The first dataset, taken from the Land Registry (2014), comprises all of the sold house prices between 1995 and 2011 in Greater London; a total of 2.2 million transactions were made during this time period. The dataset contains the following attributes: sold price, date of transaction, address, postcode unit, northings and eastings, dwelling type (terrace, detached, semi-detached and flats), new build (True or False) and tenure (leasehold or freehold). The dataset is mainly used for descriptive statistics, as it lacks primary structural attributes, such as the age and size of the property. Table 4.2 shows the house price descriptive statistics by year, where the average house price rose almost fourfold, from £110,000 to £420,000, between 1995 and 2011.

Table 4.2 The Land Registry house price dataset descriptive statistics.

Year	Count	Average	Standard Deviation	Maximum	Minimum
1995	84,640	113,610	107,766	3,750,000	50,100
1996	108,879	120,736	121,896	8,000,000	50,025
1997	133,784	132,777	139,062	7,500,000	50,015
1998	133,959	143,593	162,973	11,250,000	50,100
1999	163,078	163,128	190,815	32,477,000	50,002
2000	148,023	190,635	213,411	10,000,000	50,250
2001	162,826	205,727	222,894	24,750,000	50,020
2002	174,905	233,955	217,873	8,300,000	50,150
2003	155,299	250,084	217,197	9,250,000	50,323
2004	165,631	273,443	236,629	7,950,000	52,000
2005	137,763	289,879	266,569	15,193,950	52,000
2006	172,511	315,577	305,749	12,400,000	50,750
2007	166,569	352,442	361,652	19,000,000	52,500
2008	81,747	360,850	424,672	19,750,000	51,000
2009	75,461	361,264	385,722	12,500,000	50,000
2010	91,949	406,059	466,594	16,200,000	51,000
2011	89,809	418,594	492,858	19,250,000	50,750

The second dataset, taken from the Nationwide Building Society (2014), is the sold house price dataset for the Greater London area. The Nationwide Building Society is one of the top five mortgage providers in the UK. The dataset is, essentially, a subset of the full open-source Land Registry house price dataset and contains a total of 150,710 transactions between 1995 and 2011. The Nationwide dataset contains approximately 7% of all transactions in the Land Registry dataset. The dataset contains a wider spectrum of attributes, including: sold house price, postcode unit, date of transaction, dwelling age, dwelling floor size (square metre), number of bedrooms (0-10), dwelling type (terraces, flats, semi-detached, detached), newly built (True ,False) and tenure (freehold or leasehold). Due to this dataset's wide spectrum of attributes, it is used as the primary independent variable for the hedonic price models in the analytical chapters. Table 4.3 shows the house price descriptive statistics.

Table 4.3 The Nationwide Building Society house price descriptive statistics.

Year	Count	Average	Standard Deviation	Maximum	Minimum
1995	5,267	77,918	47,157	1,250,000	8,000
1996	12,822	84,593	51,600	1,000,000	10,500
1997	13,333	97,483	57,581	920,000	17,500
1998	12,513	118,629	71,454	935,000	21,000
1999	19,289	135,315	76,984	980,000	24,000
2000	8,998	149,837	82,712	960,000	28,500
2001	7,564	169,851	89,201	971,000	35,000
2002	9,448	197,965	92,338	998,000	42,500
2003	7,431	221,444	94,255	950,000	57,200
2004	7,613	244,226	105,900	3,340,000	65,000
2005	6,711	256,349	107,899	1,955,000	83,000
2006	10,803	273,656	131,699	3,125,000	48,125
2007	9,200	312,017	152,373	2,500,000	68,000
2008	3,965	296,064	146,106	1,625,000	94,500
2009	4,430	293,604	147,638	1,850,000	70,000
2010	5,105	319,843	171,221	1,850,000	49,950
2011	6,218	347,881	206,694	4,400,010	81,000

An important assumption in using a population sample is the consistency in the distribution of the dataset. Figure 4.18a shows the Land Registry dataset, and Figure 4.18b displays the Nationwide dataset. Table 4.4 demonstrates the quantile statistics of the two figures respectively.

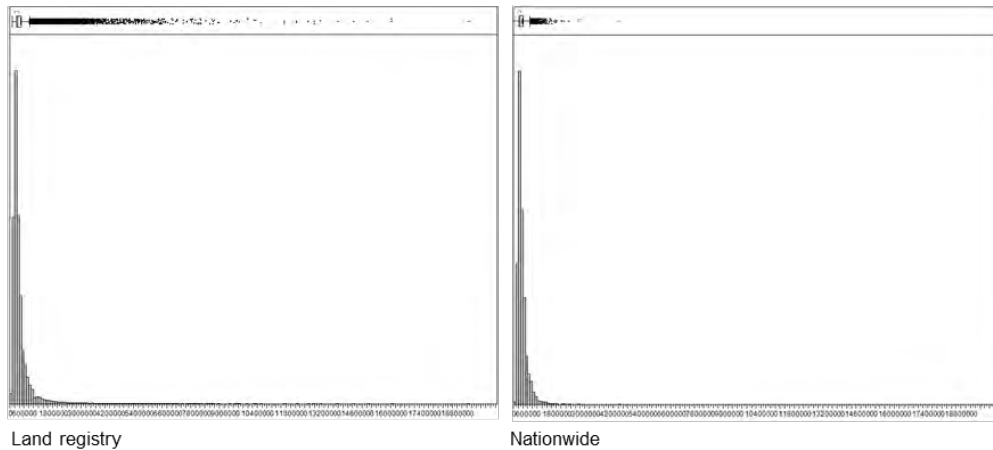


Figure 4.18 House price distribution in 2011.

- The Land Registry house price dataset.
- The Nationwide Building Society house price dataset.

Table 4.4 House Price descriptive statistics in 2011.

		Land Registry	Nationwide
maximum	100.00%	19250000	4400010
	99.50%	3171325	1300000
	97.50%	1525000	862250
	90.00%	730000	580000
quartile	75.00%	445000	410000
median	50.00%	292000	290000
quartile	25.00%	218000	227000
	10.00%	165000	180000
	2.50%	121000	140000
	0.50%	85000	110165
minimum	0.00%	50750	81000

The results show that the two datasets largely follow a long tail distribution with a similar median. The key difference between the two datasets is in the top 10% quantile range, where the Nationwide Building Society dataset is not representative. One possible reason is that the buyers denoted in this range, such as institutional, overseas and cash buyers, may not need a mortgage. The exclusion of the top quantile samples is justified for the study, as demand factors are likely to influence these types of purchases differently. Figure 4.19 shows the 2011 London house prices mapped in GIS, where red indicates higher house prices and blue indicates lower house prices. The thematic distribution in GIS is calculated using the natural break method for eight bands.

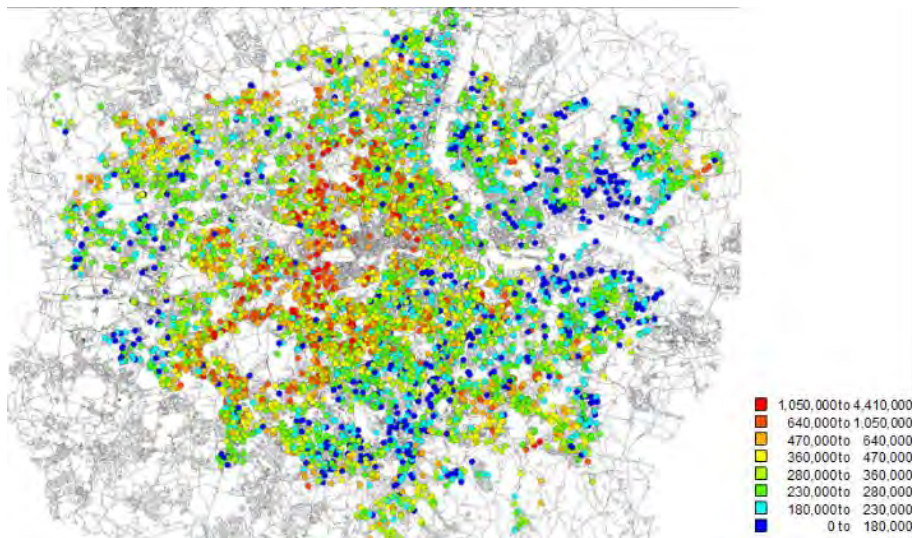


Figure 4.19 Visualisation of London House Price in 2011, from red indicating high to blue indicating low.

The figure shows the clustered nature of London's house prices with the high house price cluster starting from the top of Hampstead Heath, passing through West London and down to Richmond in the Southwest. The low house price cluster is concentrated east of Lea Valley and in the Southeast and South London. Figure 4.20 shows the change in house prices between 1995 and 2011 and demonstrates that house prices raised considerably between these time periods, though the clustering remained relatively stable.

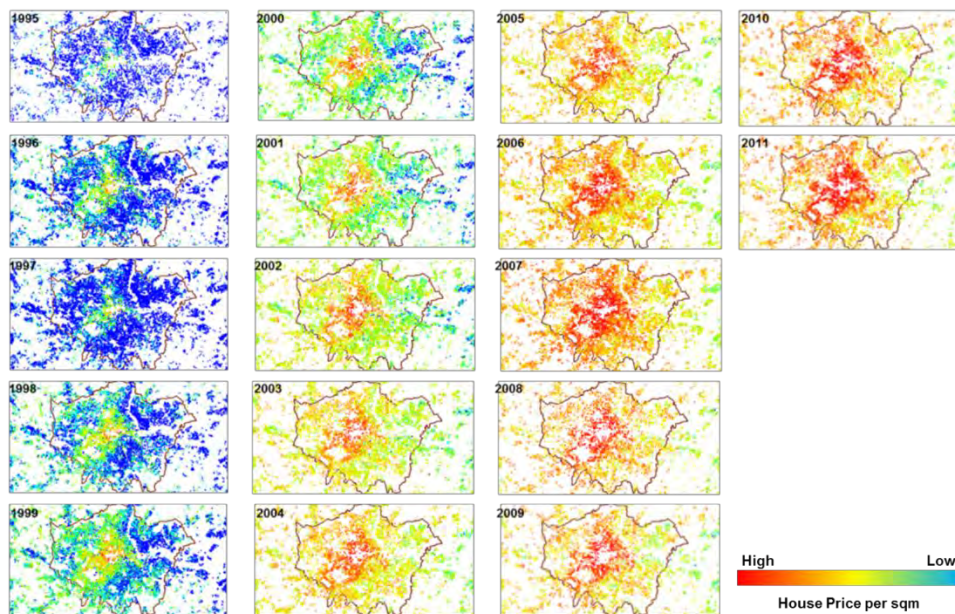


Figure 4.20 Visualisation of London house prices from 1995 to 2011, with red indicating high and blue indicating low.

Table 4.5 shows the key attributes for the property dataset, where the mean floor size is 91 square metres, the average age is 70 years old and the mean number of bedrooms is 2.41. Figure 4.21 shows there are more houses (57% of all transactions) than flats (43% of all transactions) in London.

Table 4.5 The structural attributes statistics.

Statistics	Count	Mean	Standard Deviation	Minimum	Maximum
Age	150,797	70.38	36.77	0	497
Floors	150,797	91.01	35.85	24	278
Bedrooms	150,797	2.41	0.93	1	10

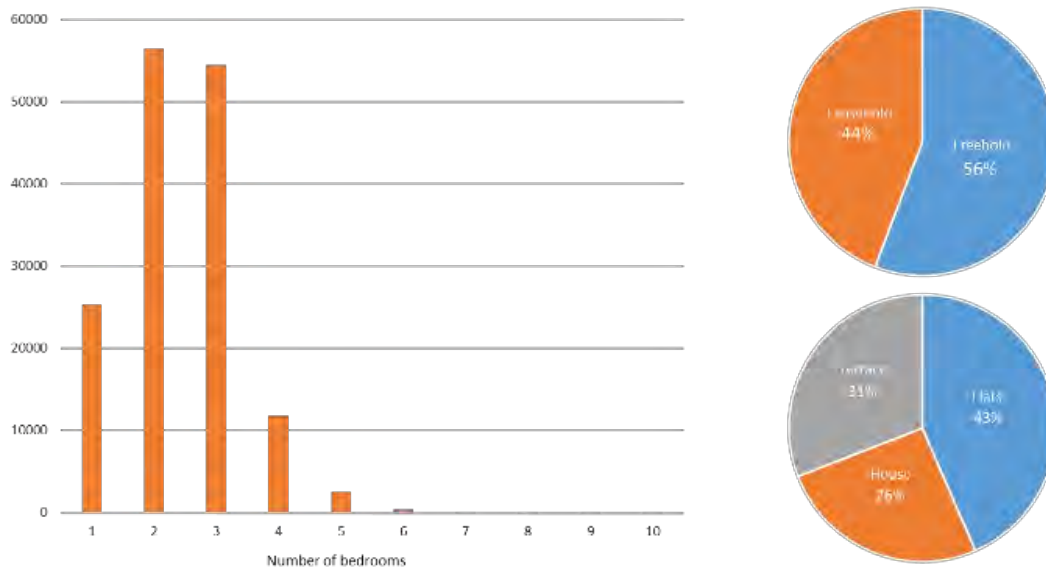


Figure 4.21 The structural attributes statistics.

- Number of bedrooms
- Tenure
- Dwelling type

Using GIS mapping, Figure 4.22 displays the spatial distribution of the six structural attributes. Listing from the top left to the bottom right, the graphs present these six attributes, which include the floor size, the number of bedrooms, the tenure (freehold or leasehold), the dwelling type (house or flat), the age of the building and the year the building was sold. The results show that greater proportions of smaller, older, leasehold flats are located in the centre, and that higher numbers of larger, newer, freehold houses are located further away from the centre.

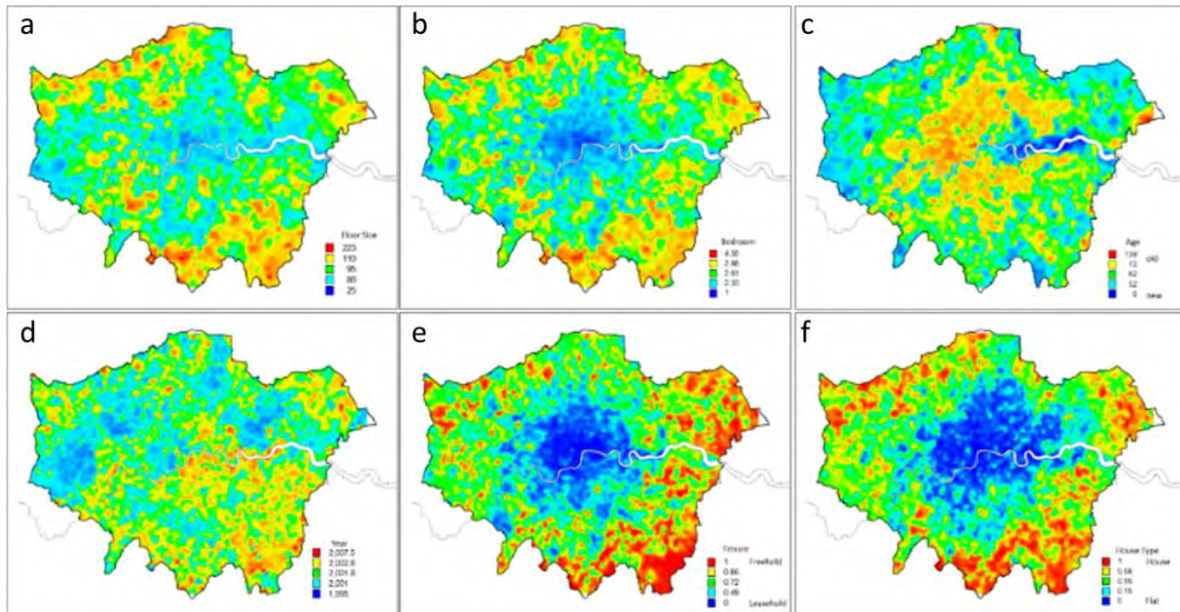


Figure 4.22 The spatial distribution of housing attributes mapped using GIS.

- a. Floor size
- b. Bedrooms
- c. Age
- d. Year of transaction
- e. Tenure (1 =Freehold, 0 = Leasehold)
- f. Dwelling type (1 = House, 0 = Flats)

4.5.1.2 The London Street Network and the Transport for London Network Datasets

The London street network dataset and the Transport for London (TfL) Tube network datasets are used to calculate both geometric and geographic accessibility measures for the analytical chapters. The London street network dataset comes from the Ordnance Survey (OS) Meridian Line dataset, which not only covers the spatial street network for Greater London but also for the entire UK (Ordnance Survey, 2014). The street network dataset has been edited manually to include some key pedestrian paths in Central London. The TfL Tube network dataset contains two sub-datasets concerning the station locations and links between the stations. The station locations dataset was taken from the TfL open data platform (TfL 2014). Based on information gathered from the TfL website, the links between the stations were constructed manually by the author as topological links. The TfL Tube network dataset includes the London Underground, the Docklands Light Railway (DLR), the London Overground and the Croydon Tramlink. The TfL Tube network excludes London's National Rail network, the London bus network, the London River Services, and the ferry network. This research recognises the exclusion of these transport modes as a research limitation.

The combined network, which the author calls the London spatial network model, has been constructed by the author following the process described in Law et al. (2011). The two networks are linked at the station, where a connection was created between the station and the street network. In London, the

spatial network dataset has a total of 113,555 segments, of which 114,018 are street segments, 535 are TfL Tube segments, and 432 are links that connect the TfL station and the street network.⁹ Figure 4.23 shows the street network dataset on the left (a) and the street and the TfL Tube network dataset on the right (b). The street network is represented by the colour black, the London Underground by the colour red, the London Overground by the colour orange and the London DLR and Croydon Tramlink by the colour cyan.



Figure 4.23 The spatial network model of London.

a. The Greater London street spatial network dataset.

b. The Greater London street and TfL network dataset.

Four spatial network models concerning the years 1995 to 2011 have been built for the analytical chapter. Figure 4.24 shows these four models and the time period in which each of the public transit lines was constructed. The four networks include the Jubilee Line Extension and the DLR second stage extension in 1999, the Croydon Tramlink in 2000, the DLR third stage extension in 2004, the DLR Stratford extension in 2009 and 2010 and the East London Overground line in 2010. Table 4.6 shows the tabulation for these four spatial network models.

Table 4.6 Street and TfL spatial network datasets.

	1995	2000	2005	2011
TfL Station-link	432	432	432	432
TfL Network	463	520	524	535
Street	114,018	114,018	114,018	114,018
			Total Segments	114,985

⁹ Various GIS processes were applied for preparation of spatial network analysis

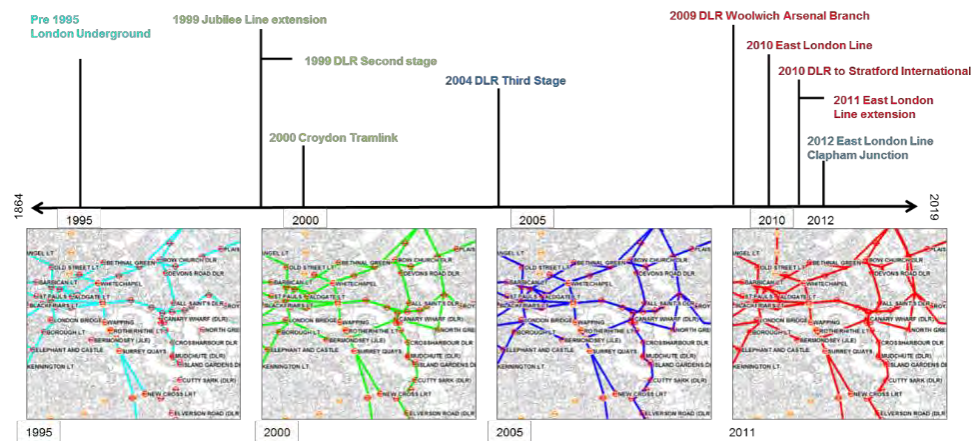


Figure 4.24 Greater London Public Transport Innovation projects between 1995 and 2011.

This multi-layered spatial network model is used to calculate various geographic and geometric accessibility measures for the three analytical chapters. Figure 4.25 illustrates one such measure, the closeness centrality of London, where red is higher accessibility and green is lower accessibility.¹⁰ These measures are used as the key variables in the first analytical chapter and as control variables in the second and third analytical chapters.

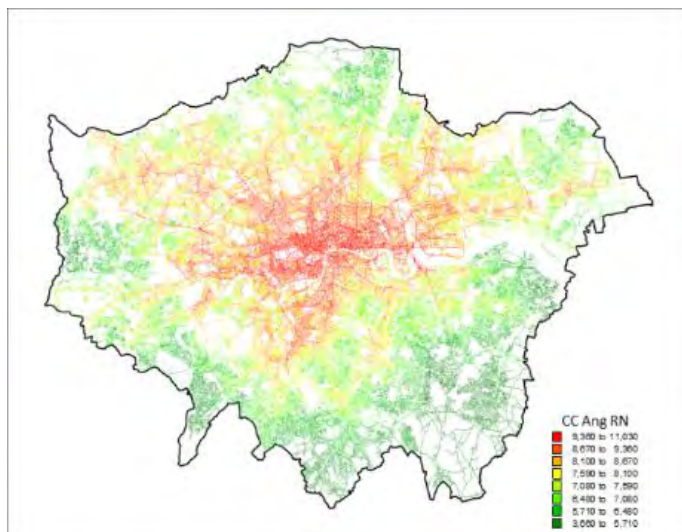


Figure 4.25 The Greater London closeness centrality with a radius of 20 kilometres.

4.5.1.3 The London Retail Address Dataset

The retail address points, taken from the Valuation Office Agency address dataset, are used to calculate the retail amenity measure. In 2010, there were 98,078 retail address points in London, the majority of which were shops, followed by restaurants, cafés, salons, showrooms, kiosks, banks and betting shops.

3 Spatial network closeness centrality or integration in the spatial configuration literature measures the reciprocal sum of the shortest path between every origins (i) to every destinations (j). (Sabadussi, 2005 as mentioned in Iida and Hillier, 2005)

As the dataset includes 600 categories, many of these were subsequently grouped; these groups are tabulated in Table 4.7.

Table 4.7 The retail address points dataset.

Land-use Type	Count
Shop and Premises	77,776
Restaurant and Premises	6,552
Café and Premises	2,037
Hairdressing Salon and Premises	1,588
Showroom and Premises	1,412
Kiosk and Premises	1,271
Bank and Premises	1,098
Betting Shop and Premises	1,028
Other	5,316
Total	98,078

Figure 4.26 shows the spatial density distribution of the retail addresses, which was visualised in GIS by using a 400-by-400 metre grid. In this figure, the red cells denote a higher density of retail amenities, and the green cells represent a lower density of retail amenities. The result shows a clustering of retail uses in the centre of the city, which then disperses into smaller centres.

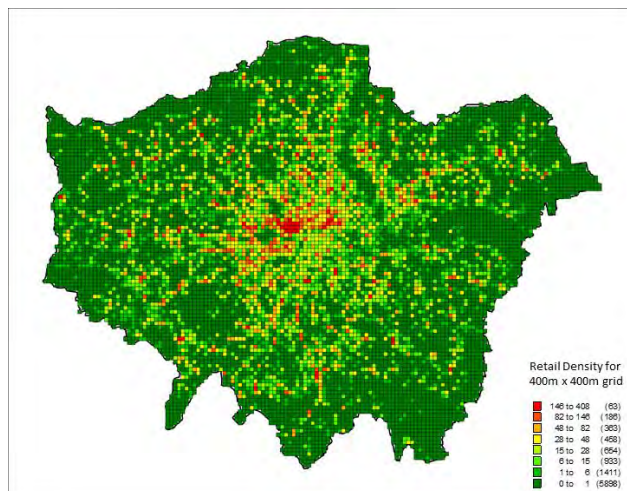


Figure 4.26 Visualisation of the retail amenity using a 400-by-400 metre grid, where red denotes a higher density of retail and green denotes a lower density of retail.

The London Retail dataset is used to calculate the retail amenity accessibility within 800 metres (ra_i) for the analytical chapter. The measure is a type of cumulative opportunities measure in accessibility literature.¹¹ In spatial configuration literature, the measure is synonymous to the place syntax measures of accessibility (Marcus, 2000). This research uses retail accessibility as a control variable.

¹¹ $ra_i(r800m) = \sum retail_j(r800m)$

4.5.1.4 The UK School Performance Dataset

The UK school performance dataset for 2011, taken from the Department of Education website (2015a), is used to calculate school amenity measures for the hedonic price model. The inclusion of the variable is to continue from previous research on the importance of school quality in residential location choice (Gibbons 2005). There are 3,138 schools in the Greater London area. Of these, there are 2,474 primary schools, 933 secondary schools and 721 sixth form schools. Table 4.8 shows the tabulation of this dataset.

Table 4.8 Secondary school dataset (DOE, 2015a).

Local Authority	All	Primary	Secondary	16-18
Barking and Dagenham	65	51	18	18
Barnet	158	119	49	42
Bexley	91	73	22	18
Brent	101	82	28	25
Bromley	139	115	37	32
Camden	96	76	30	21
City of London	5	5	3	2
Croydon	149	114	44	39
Ealing	109	91	33	28
Enfield	106	86	32	29
Greenwich	100	74	29	19
Hackney	115	92	43	24
Hammersmith and Fulham	80	62	27	18
Haringey	98	81	23	15
Harrow	78	56	23	24
Havering	93	72	26	10
Hillingdon	111	86	32	30
Hounslow	90	68	29	23
Islington	70	56	17	15
Kensington and Chelsea	81	62	32	22
Kingston upon Thames	63	47	19	17
Lambeth	98	74	26	22
Lewisham	98	85	23	16
Merton	73	60	20	13
Newham	104	84	29	12
Redbridge	95	77	27	21
Richmond upon Thames	81	68	27	23
Southwark	113	87	32	24
Sutton	71	54	22	20
Tower Hamlets	118	86	35	26
Waltham Forest	88	68	26	20
Wandsworth	114	96	34	27
Westminster	87	67	36	26
Total Number of Schools	3,138	2,474	933	721

For this research, the average score per student, calculated by using each of the Key Stage examinations, operates as an indicator of the school's performance (DOE, 2015b). This includes Key Stage 5 (KS5 A-levels) and Key Stage 4 (KS4-GCSE) for secondary school and Key Stage 2 (KS2) for primary school.

Table 4.9 School average scores in 2011 (DOE, 2015a).

2011 School Performance	Count of Average Point Score per Student	Average of Average Point Score per Student	Standard Deviation of Average Point Score per Student	Maximum of Average Point Score per Student	Minimum of Average Point Score per Student
KS5	428	737	162	1320	293
KS4	641	411	156	781	2
KS2	1659	27	3	33	15

Figure 4.27 shows the spatial distribution of KS2 (yellow), KS4 (orange) and KS5 (red) schools at the top; the respective average scores are visualised directly below, where red denotes a higher average score and blue denotes a lower average score.

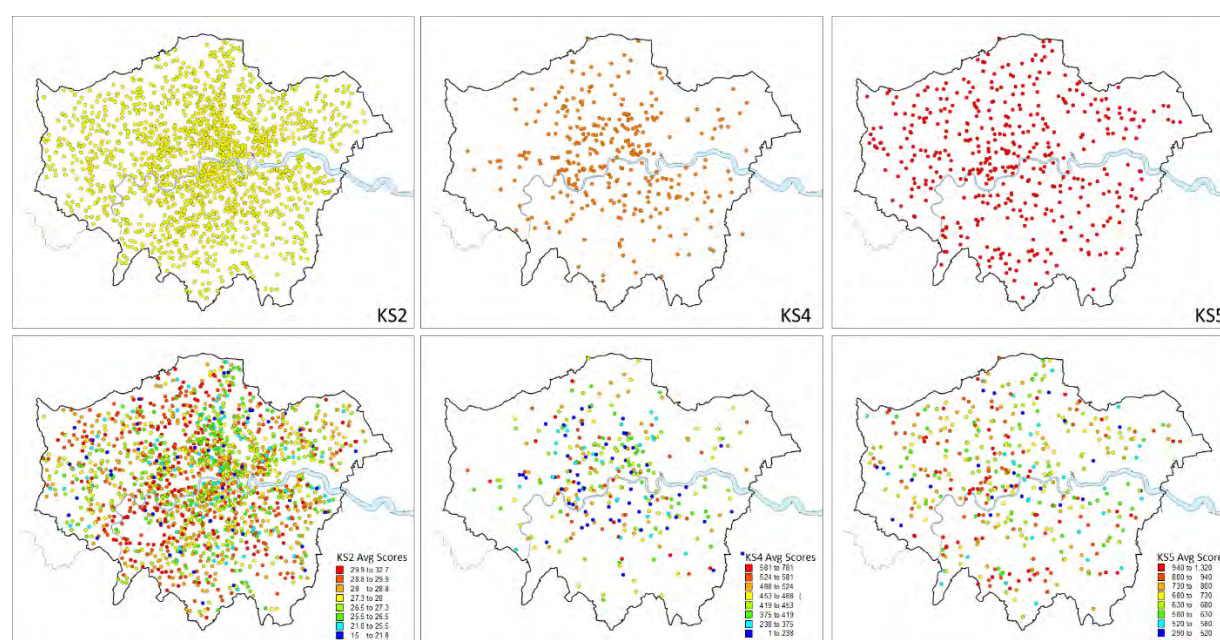


Figure 4.27 Visualisation of Greater London secondary school average A-level scores, where red denotes higher scores and green lower scores (DOE, 2015a).

- KS2 locations
- KS4 locations
- KS5 locations
- KS2 average scores
- KS4 average scores
- KS5 averages scores

The General Certificate of Education Advanced Level, also known as A-Level, is used as the school performance parameter in the analytical study. The key reason is that the A-level (KS5) average scores achieve a greater correlation than both KS4 and KS2 average scores in the hedonic price model. The dataset is used to calculate the average school score within 800m (SA_{ij}) of the property for the analytical chapter. The school amenity measure is a type of cumulative opportunities measure in the accessibility literature¹². In this study, school performance amenity is used as a control parameter.

4.5.1.5 The London Heritage Parks and Gardens Dataset

The London Heritage Parks and Gardens dataset is used to calculate park amenities for the analytical chapters (English Heritage, 2014). The dataset of registered parks and gardens comes from the English National Heritage website and includes a range of planned open spaces, such as public parks, cemeteries, private grounds, and town squares; to indicate a space's level of significance, it is assigned one of three grades. The dataset of parks and gardens includes 1,640 entries in the UK, of which 150 are located in Greater London. Also comprised in the dataset are large Royal Parks such as Regent's Park, Hyde Park, Richmond Park and St. James's Park, and large neighbourhood parks such as Dulwich Park and Walpole Park. Figure 4.28 shows the location of these parks and gardens; a full list the included parks and gardens can be found in Appendix B.

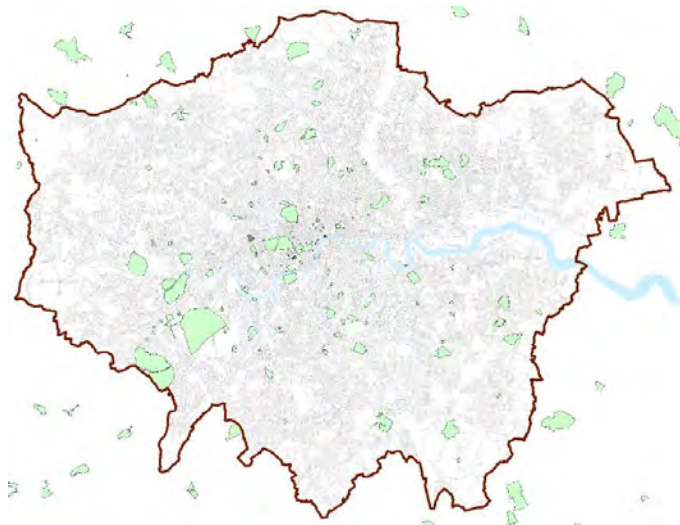


Figure 4.28 Greater London registered parks and gardens (English Heritage, 2014).

For the empirical study, this dataset is used to calculate minimum walking distances to parks and gardens (P_{ij}), which is a type of spatial separation measure in accessibility literature.¹³ In this study, accessibility to parks and gardens is used as a control variable.

¹² $SA(R800m)_i = \frac{\sum A \text{ Level Score}_j}{n}$

¹³ $P_i = \min d_{ij}$

4.5.1.6 UK Census Employment Dataset

The UK Census employment dataset aggregated at the LSOA level is used to calculate the gravitational potential for the first research strand (ONS, 2015).¹⁴ There are 4,765 LSOAs in Greater London with employment statistics for the years 1998,¹⁵ 2000, 2005 and 2011. Table 4.10 shows the tabulation of this dataset, and Figure 4.29 exhibits the dataset's spatial distribution.

Table 4.10 Job statistics between 1998 and 2011 at the LSOA level.

	Jobs in 1998	Jobs in 2000	Jobs in 2005	Jobs in 2011
Mean	754	818	864	911
Std.Dev.	4,039	4,475	4,545	5,291
Maximum	226,801	250,478	250,229	299,221
Minimum	9	9	9	9
Total	3,594,163	3,895,464	4,117,870	4,340,675

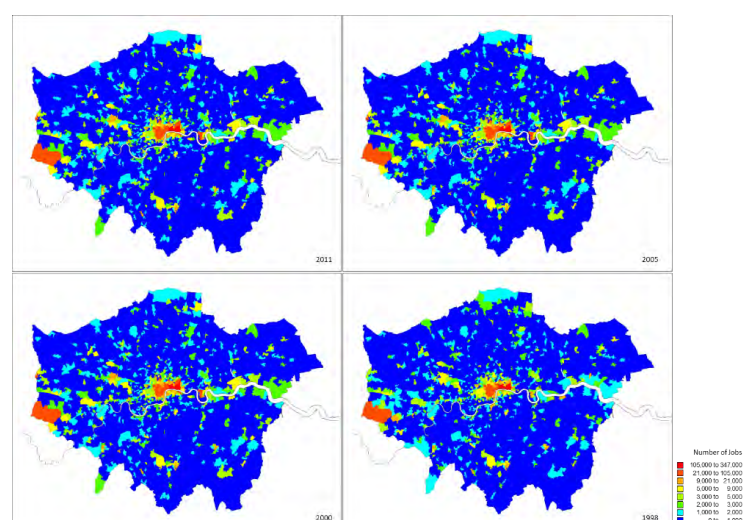


Figure 4.29 Visualisation of the London jobs spatial distribution from 1998 to 2011, where red indicates a more jobs and blue indicates fewer jobs.

4.5.2 Research Strand Two Datasets

Five datasets for the second research strand are described in this section, including the spatial street network dataset, the Statistics for Wards dataset, the Medium Super Output Area (MSOA) dataset, the Lower Super Output Area (LSOA) dataset and the Postcode Unit dataset. Figure 4.30 shows these five datasets overlaid on top of the Thamesmead area in London.

¹⁴ The Middle Super Output Area contains approximately 2000 to 6000 households in England. The Lower Super Output Area contains 1000 to 3000 households in England. The average population of an MSA in London in 2010 was 8,346, compared with 1,722 for an LSOA and 13,078 for a ward.

¹⁵ Employment statistics for 1995 are not available at this granularity.



Figure 4.30 The UK Postcode Unit, LSOA, MSOA, Ward and GLA visualised for Thamesmead in London.

4.5.2.1 The Spatial Street Network Dataset

The GLA spatial street network dataset is used to calculate the St-LA in the second research strand and the St-HS in the third research strand. The spatial street network dataset comes from the OS Meridian Line dataset as a basis, in which the pedestrian paths have been manually added (Ordnance Survey, 2014). In London, the spatial network dataset has a total of 113,555 segments, as shown in Figure 4.31.

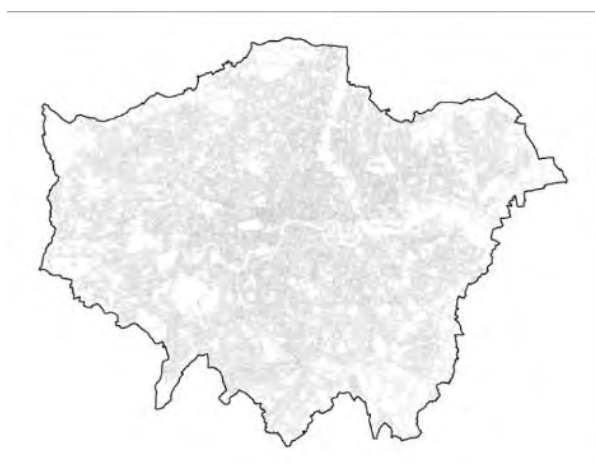


Figure 4.31 Greater London Spatial Street Network dataset.

Applying the modularity optimisation method, as described in Chapter 3, to the spatial street network creates five levels of nested St-LA partitions. Table 4.11 summarises the results for these divisions. The first separation has 10,470 local areas, the second segment has 2,838, the third partition has 686, the fourth section has 207 and the fifth separation has 167.

Table 4.11 Street-based local area statistics.

Street-based Local Area	1	2	3	4	5
Count	10,470	2,838	686	207	167
Mean	11	40	166	549	680
Standard Deviation	5.5	24	109	257	268
Maximum	67	202	765	1,243	1,572
Minimum	4	8	20	73	102

Figure 4.32 shows the Greater London St-LA mapped in GIS, where the different colours correspond to various community memberships. This partition is selected for the analytical chapters, as it achieves a better correlation with house prices. Visually, the results show a clear distinction between local areas separated by the River Thames, such as the Isle of Dogs and the Royal Docks; local areas divided by railway tracks, such as the Crouch End and Harringay; and local areas parted by the Lea Valley. The local area boundaries in Central London are not as apparent, due to the porosity of the street grid in the city centre.

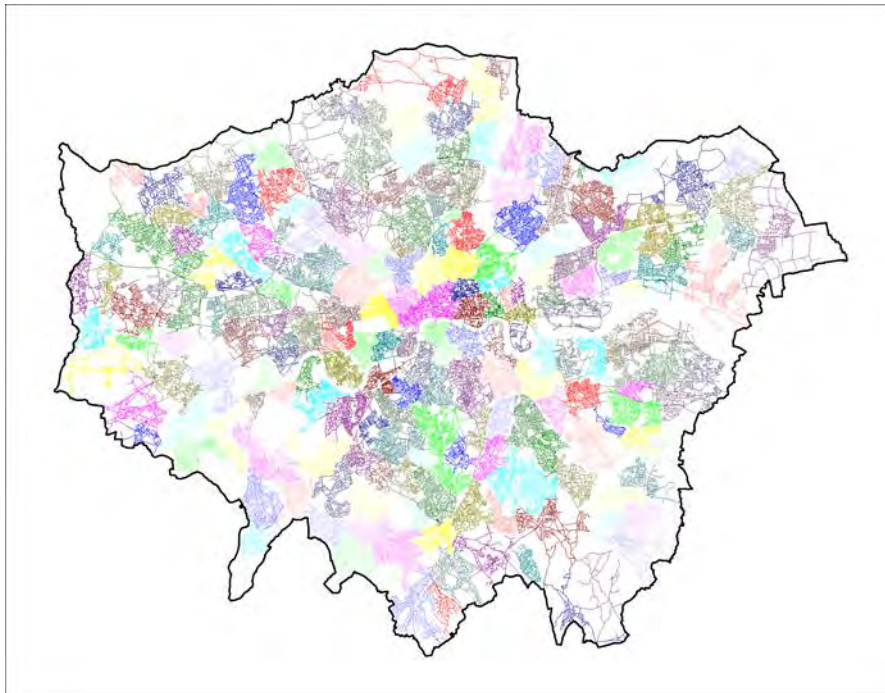


Figure 4.32 Visualisation of St-LAs in the Greater London area.

4.5.2.2 Postcode Unit Dataset

The Postcode Unit dataset is used to calculate the Postcode Unit local area in Research Strand Two and the Postcode Unit housing submarket in Research Strand Three. This geography is the smallest in the UK after individual addresses and was defined by the Royal Mail to identify postal delivery areas (ONS, 2015). There are approximately 3.2 million postal addresses in the UK and 197,066 Postcode Units in Greater London. Figure 4.33 illustrates the postcode unit boundaries for Greater London, and Table 4.12 exhibits the descriptive statistics for Greater London.

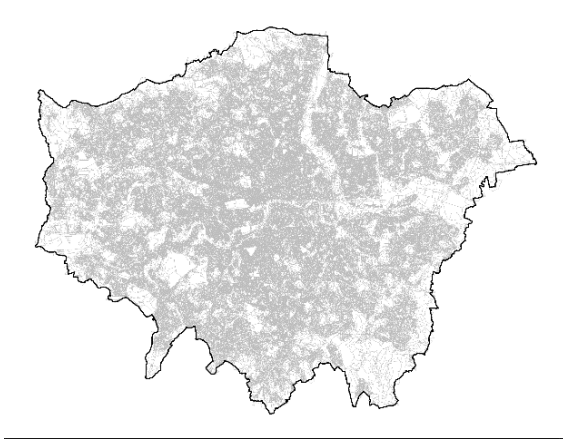


Figure 4.33 Postcode Units in Greater London.

Table 4.12 Postcode Unit statistics for Greater London.

Postcodes	197,065
Addresses	3,261,086
Average	16.44
Std. Dev.	16.55
Maximum	100
Minimum	1

4.5.2.2 Census Area Statistic (CAS) Ward

The ONS Census Area Statistical (CAS) Ward dataset is used to calculate the Ward local area in Research Strand Two and the Ward housing submarket in Research Strand Three. The geography was defined by the ONS as the key administrative geography in the UK for the election of local government officials.¹⁶ There are 651 statistical wards in Greater London, with an average of 11,000 residents and an average size of 245 hectares per electoral ward. Figure 4.34 illustrates the statistical ward boundaries for Greater London, and Table 4.13 displays the descriptive statistics for this geography.

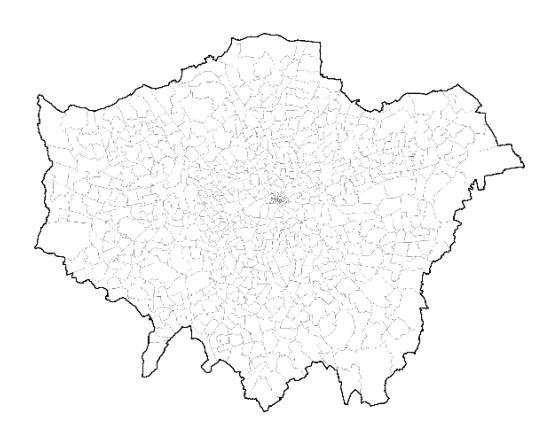


Figure 4.34 The UK Electoral and Census Area Statistics Ward.

¹⁶ Districts, metropolitan boroughs, unitary authorities and London boroughs in England. Unitary Authorities in Wales. Council areas in Scotland. District council areas in Northern Ireland.

Table 4.13 The electoral ward statistics for population and employment.

	Ward	Mean	Standard Deviation	Maximum	Minimum
Hectares	651	245.3	257.3	2,903.3	0.3
Population	651	11,038.8	5,028.0	25,297.0	0.0
Employment	651	5,851.4	16,687.2	312,171.0	0.0

4.5.2.3 Super Output Area (LSOA, MSOA)

The ONS Super Output Area dataset is used to calculate the LSOA and MSOA local areas for Research Strand Two and the LSOA and MSOA housing submarkets for Research Strand Three ¹⁷ (ONS, 2015). Output Area (OA) units are produced by the ONS, according to the multiple criteria of size, shape, natural boundaries, population, tenure and dwelling type. The OAs are then aggregated to form LSOAs and MSOAs. There are 983 MSOAs in Greater London with a mean population of 7,200 and an average size of 162 hectares. There are 4,765 LSOAs in Greater London with a mean population of 1,500 and an average size of 33.5 hectares. Figure 4.35 shows the spatial distribution of MSOAs (a) and LSOAs (b) for Greater London. Tables 4.14 and 4.15 summarise the descriptive statistics for this geography.

Table 4.14 The MSOA population and employment statistics (ONS, 2015).

	MSOA	Mean	Standard Deviation	Maximum	Minimum
Hectares	983	162.2	188.4	2,243.0	29.4
Population	983	7,296.2	1,109.6	12,361.0	5,021.0
Employment	983	3,871.6	12,660.6	312,171.0	500.0

Table 4.15 The LSOA population and employment statistics (ONS, 2015).

	LSOA	Mean	Standard Deviation	Maximum	Minimum
Hectares	4,765	33.5	64.4	1,579.7	1.8
Population	4,765	1,536.7	129.3	2,686.0	913.0
Employment	4,765	834.4	4,461.5	244,926.0	9.0

¹⁷ The middle super output area contains approximately 2000 to 6000 households in England. The lower super output area contains 1000 to 3000 households in England. The average population of an MSOA in London in 2010 was 8,346, compared with 1,722 for an LSOA and 13,078 for a ward.



Figure 4.35 Super Output Area of Greater London.

- a. The MSA mapped for Greater London.
- b. The LSOA mapped for Greater London.

4.5.3 Research Strand Three Dataset

Research Strand Three does not use any new datasets. Instead, this research strand combines the datasets from the first two in order to create six types of housing submarket variables; these six submarkets are shown in Figure 4.36. Details of these six specifications are discussed in the third analytical chapter.

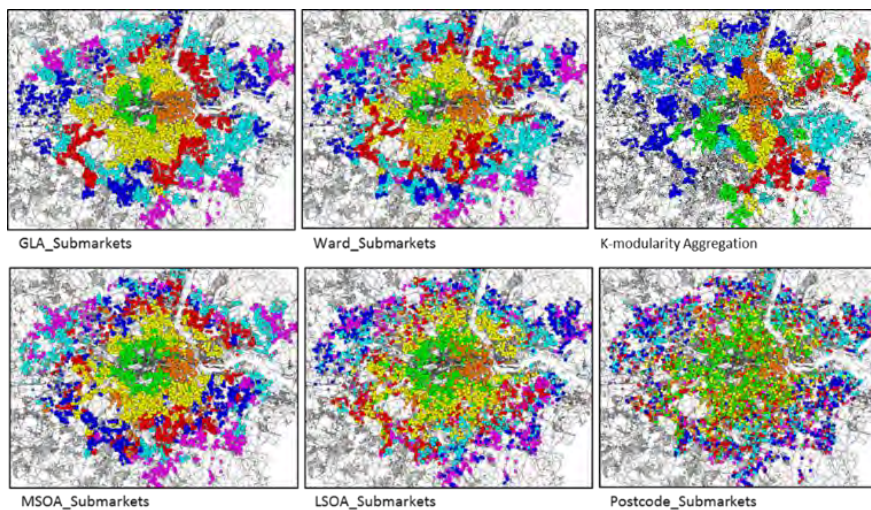


Figure 4.36 Different housing submarkets used for Research Strand Three.

4.5.4 Discussion

This section has described all of the datasets used for the research, including the house price datasets, the spatial network dataset, the local amenity dataset and the local area datasets. These datasets are used to calculate geographic and geometric accessibility variables, St-LA variables and St-HS variables that are utilised in the thesis. The next chapter is the first research strand, which will examine the extent to which geometric accessibility measures associate with house prices in London and the degree to

which geometric accessibility compares with geographic accessibility measures.

Key Data Sources for the chapter:

Contains Ordnance Survey data © Crown copyright and database right 2013 and 2015

Contains National Statistics data © Crown copyright and database right 2015

Contains data produced by Land Registry © Crown copyright 2015

Data provided by the Valuation Office Agency contains public sector information licensed under the Open Government Licence v1.0.

Chapter 5

Spatial Network Accessibility Effects on House Prices

5.1 Introduction

Estimating the effects of employment accessibility on house prices using the hedonic price approach is an extensively examined topic in housing studies (Adair et al., 2000). Empirically, accessibility effects on house prices have been traditionally estimated using either the distance to the central business district (CBD) measure (Alonso, 1964) or the gravitational potential to employment measure (Hansen, 1949). More recently, geometric accessibility measures (Jiang et al., 2002), were found to be significant (Xiao et al., 2015; Law et al., 2013; Chiaradia et al., 2012). Despite its significance, empirical research on applying geometric accessibility measures is limited. There are two aims for this analytical chapter. The first is to study the extent to which centrality measures can be used to measure the impact that transport projects have on house prices in London; the second is to compare the extent to which geographic and geometric accessibility measures are significant in associating with house prices. Deconstructing accessibility effects provides evidence to inform policies on transport and urban design projects. This study employs a hedonic price approach to estimate the effects of both geographic and geometric accessibility on transport projects implemented in London between 1995 and 2011. The study focuses on major intra-city London public transport projects, such as the London Underground Jubilee Line Extension (JLE), the East London Overground (OG) Line, the Docklands Light Railway (DLR) and the Croydon Tramlink (CRTL). The remainder of this chapter is organised as follows. Section 1 introduces the chapter. Section 2 describes the three accessibility measures for the analytical work. Section 3 presents the hedonic price modelling framework used to answer the research question. Section 4 describes the case study and the datasets. Section 5 reports the empirical results. Section 6 provides a general discussion of the key findings and limitations of the research. Figure 5.1 illustrates the focus of this chapter, which concerns the property-level accessibility effects on house prices within the framework as introduced in Chapter 4.

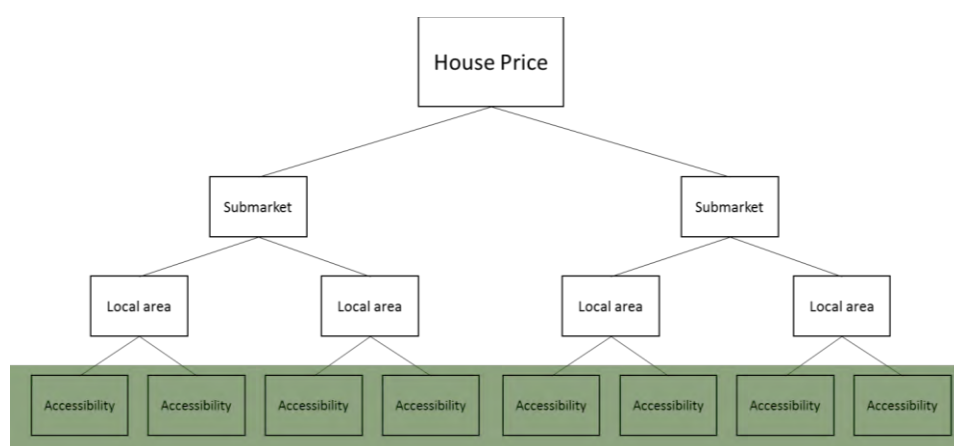


Figure 5.1 This chapter focuses on the accessibility effects on house prices.

5.1.1 Strand One Research Question Definition

As described in Chapter 2, the theory of spatial equilibrium (Glaeser, 2008) and the bid-rent model (Von Thunen 1826; Alonso, 1964) form the primary mechanisms to explain intra-city house price variations. For example, people are willing to live further away from their job if the rent per square metre is lower and in a smaller space if their commuting cost is reduced. The essence of the monocentric model is its simplicity in explaining land rent through accessibility. Adapted from the monocentric model, distance to the CBD has become the most widely used location variable in hedonic price models (Sirmans et al., 2006; Kain and Quigley, 1970). This measure of spatial separation assumes a homogeneous house price gradient that uniformly declines from a central point of employment. More recent models acknowledge the complex structure of cities, resulting in the use of access to multiple employment centres (Ahfeldt, 2011; McDonald and McMillen, 1990; Orford, 1999). The motivation is to move away from all economic activity being concentrated in a single point and towards a polycentric employment distribution. Previous literature has consistently shown observable geographic accessibility effects on house prices in London using both the distance to the CBD and gravity-type measures (Ahfeldt, 2011; Gibbons and Machin, 2005). Specifically, a study currently being conducted by Hayman and the author has found that the most popular accessibility measures in the hedonic price model in descending order are: the spatial separation measure, the cumulative opportunities measure and the gravity-based measure. The research has also found the most popular types of distance in defining accessibility are, in order: Euclidean distance, metric distance and travel time. There is a clear absence of geometric accessibility measures from hedonic price models and a lack of comparative research across accessibility measures.

One strand of research that consistently examines the geometric accessibility effects on house prices is space syntax literature. Empirically, early space syntax studies found geometric accessibility measures associate positively with office rental data in Berlin (Desyllas, 1997). More recent empirical research focused on the relationship between spatial configuration and council tax bands in London (Chiaradia, Hillier, Barnes and Schwander, 2012a; 2012b). These studies demonstrated the significant positive correlation between geometric accessibility and economic performance. In a conference proceedings paper, Law et al. (2013) showed the significant relationship between geometric accessibility and house prices in London using an ordinary least squares (OLS) specification. Xiao et al. (2015) similarly presented how urban configuration was associated with house prices in Cardiff (Webster, 2010). Despite the identification of the correlations between geometric accessibility measures and house prices, these researchers did not explicitly study the impact that transport projects have on house prices; secondly, these studies did not specifically compare geographic and geometric accessibility measures. This research, therefore, asks the following question:

Research strand one question: To what extent are measures of geometric accessibility associated with intra-city house price variations? how does geometric accessibility compare with geographic accessibility measures when associating with house price?

5.2 Geometric Accessibility Framework

This research suggests that the association between accessibility and house prices can be deconstructed into geometric accessibility effects and geographic accessibility effects. Figure 5.2 conceptualises the effects of these two accessibility measures; there are both overlapping effects (represented by the middle intersecting region) and separate effects on house price (signified by the blue and the orange sections that are not overlapping). This research conjectures that, in the hedonic price model, both geographic accessibility and geometric accessibility are significant when associated with house prices.

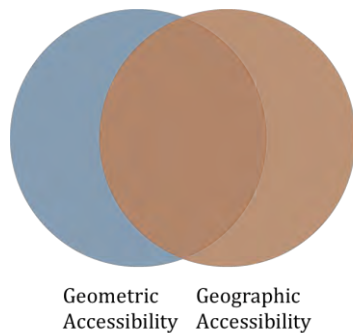


Figure 5.2 Diagram showing how geometric accessibility and geographic accessibility overlap with each other.

To explain how geographic and geometric accessibility influence house prices is beyond the scope of this investigation. Rather, this research studies the extent to which the two effects exist when associated with house prices. Possible mechanisms used to determine how geometric accessibility relates to house prices can come from different sources. From one perspective, geometric accessibility can capture the employment effects from the geographic accessibility measures. This can be drawn from the movement economy theory (Hillier et al 1996) where spatial configuration is a determinant of commercial land use via the “natural movement” it generates. This spatial process in turn generates the “destination-movement” of geographic accessibility model. From another perspective, geometric accessibility may capture the non-employment accessibility effect, which can be represented by the non-overlapping elements in Figure 2. This effect can be interpreted as generic accessibility effects (Webster 2010) or a network effect generated by the grid (Jacobs, 1961). This effect will be further discussed in the final section of this chapter. The next section describes the geographic and geometric accessibility measures used for this study.

5.2.1 Accessibility Specification

As mentioned in Chapter 3, accessibility is defined as a measure of potential interactions, or relative proximity or nearness, in an environment for individuals or spaces open to everyone (Hansen, 1949). These measures can be separated into two broad classes, namely geographical accessibility and geometric accessibility (Jiang et al. 2002). Geographical accessibility is a function of the destination attraction and the impedance or distance between the origin and the destination. Geometric accessibility focuses on the spatial network itself and the specification of different distances and radii.

Figure 5.3 conceptualises the key difference between the two accessibility measures, where geographic accessibility concerns access to specific destinations, such as those related to employment, whilst geometric accessibility concerns access to all spaces. One geographic accessibility measure and two geometric accessibility measures are applied in the empirical section.

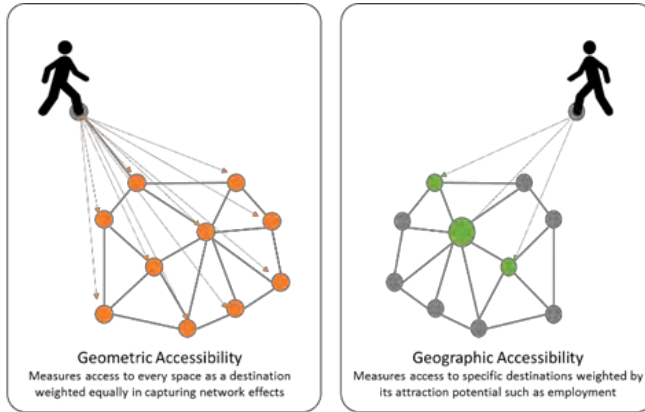


Figure 5.3 Geographic and geometric accessibility.

The first measure, analogous to Newtonian physics, is the geographic accessibility measure of the gravitational potential to employment (Hansen, 1949). Accessibility, in this instance, is a function that is directly proportional to employment size and inversely proportional to the distance between the household location and employment location. Accessibility is summed up for each employment region. This research does not imply any functional form for the measure and keeps the parameters to unity for simplicity reasons, such as making geographic and geometric accessibility more comparable. For future research, differences in the functional form can be easily adopted for both measures.

$$GP_i = \sum \frac{o_j(r)}{d_{ij}(r)}$$

Where

GP_i is the accessibility at i

o is the attraction at j

d_{ij} is a measure of impedance between i and j

r is the radius threshold

Equation 5.1

The second is the space syntax measure of geometric accessibility, also known in network science as integration or closeness centrality (Hillier and Hanson, 1984). Space syntax calculates the reciprocal average shortest path length between every origin (i) to every destination (j). More simply, it measures the 'to movement' potential to reach all of the nodes in a network (Hillier and Iida, 2005; Freeman, 1977; Sabidussi, 1966). Empirically, closeness centrality has been found to associate positively with residential property values and commercial rent.

$$CC_i(r) = \frac{\sum n_i(r)}{\sum d_{ij}(r)}$$

where

n is the total number of nodes reachable from i at the network up to metric radius r

CC is the closeness centrality at i up to radius r

d_{ij} is a measure of impedance between i and j

r is the radius threshold

Equation 5.2

From graph theory, the third measure is known as harmonic centrality (Boldi et al., 2014), which applies an inverse distance function to each node. This measure interprets the attraction in geographic accessibility as the density of the nodes. Therefore, being closer to more nodes brings higher accessibility. The spatial network is clearly central to this measure; as a result, the measure is still considered a geometric accessibility measure.

$$HC_i(r) = \sum \frac{1}{d_{ij}(r)}$$

where

HC_i is the accessibility at i

d_{ij} is a measure of impedance between i and j

r is the radius threshold

Equation 5.3

These three measures capture the positive accessibility effects on house prices in a hedonic model. Choice or betweenness centrality in graph theory (Hillier and Iida, 2005), which correlates positively with movement potential and negatively to house prices, is a different type of geometric measure that captures through-movement potential effects in space syntax. It encapsulates the negative accessibility effects on house prices from pollution and noise in a hedonic price model (Penn et al., 1997; 1998). For this study, this measure is only used as a control variable. An important topic in space syntax literature is the comparison between different distance types, such as angular, topological, travel time and metric distance. This study focuses on comparing the effects of travel time and angular distance. Another important topic, often discussed in space syntax literature, is the comparison between different radii. Radius is defined as the cut-off threshold for measuring accessibility. For example, closeness centrality **CC** at radius **X** measures segment **i**, its inverse average distance to all destinations **j**, and up to radius **X**. This study focuses on testing the effects of radius infinity and radius 60 minutes on measuring both global geographic and geometric accessibility. Radius 60 minutes was chosen to reflect global accessibility with a travel time budget, as the average commuting time in Southeast England is somewhere between 50 to 70 minutes. Future research will test a range of radii to examine local effects on house prices. Due to the confounding effects which occur between radii, this research strand mainly focuses on the global measures. The use of the principal component analysis can reduce the problem of collinearity. However, it makes the estimates of the accessibility variable much less interpretable. The comparison between travel time and angular distance and the use of a travel time radius are novelties in space

syntax research. Table 5.1 shows the parameter space to be tested for this chapter. The next section will provide the empirical framework for testing the different accessibility measures. Please see Chapter 3 for a more thorough discussion on the accessibility measures and its specifications.

Table 5.1 The parameter space for the three accessibility measures.

	Angular	Travel Time	Angular	Travel Time
	Radius Infinity		Radius 60 Minutes	
Gravitational Potential (GPi)				
Harmonic Centrality (HCi)				
Closeness Centrality (CCi)				

5.3 Empirical Method

Rosen (1974) described the hedonic price approach as where a differentiated product, such as housing, is made up of ‘utility-bearing’ characteristics. Within the hedonic price modelling framework, a household maximises its utility by moving along their hedonic price function until reaching a point where the demand for the amenity (marginal willingness to pay) is equal to the supply of the amenity (marginal willingness to offer). This model allows house prices to be broken down into a bundle of utility-bearing characteristics, including structural characteristics, such as size and age, neighbourhood amenities, such as school quality and shops, and location accessibility, such as the accessibility to employment (Sirmans et al., 2006; Cheshire and Sheppard, 1995). To answer the research question, this study adopted the hedonic price approach.

5.3.1 Fixed-Effect Hedonic Price Approach

In order to compare the three different measures, this study estimated the accessibility effects on house prices in London by using the fixed-effect (FE) hedonic price regression model introduced by Gibbons and Machin (2005). This model is argued as being more robust and accurate in approximating the effects of accessibility than the traditional OLS regression model and other types of spatial econometric regression models (Gibbons and Overman, 2012). The following section more formally describes the FE hedonic price approach. The starting point for the FE approach is the standard OLS regression model; this model is described by the following equation.

$$\text{Log}(P_{it}) = \beta_1 X_{it} + \beta_2 \text{Log}(A_{it}) + f_i + g_t + e_{it}$$

where

P_{it} is the price of a property for i = postcode

X_{it} represents a vector of controlled independent variable

A_{it} is the accessibility variable in the model

f_i represents place-specific unobserved variance

g_t represents the general time dummy

e_{it} is the error term

Equation 5.4

In Equation 4, the house prices for each postcode¹⁸ unit were regressed against a vector of dwelling-specific, location-specific and neighbourhood-specific variables, denoted by \mathbf{X} ; e represented the error term, g designated the general time trend and f signified the unexplained variance in the model. The first order condition of this function was the implicit price for the attribute. Table 5.2 shows a list of the variables used for this hedonic price model (Ahlfeldt, 2011; Smith, 2010; Gibbons and Machin, 2008; Sirmans et al., 2006; Gibbons and Machin, 2005; Cheshire and Sheppard, 1995).

Table 5.2 Specification of the hedonic model variables.

Name	Type	Data
Gravitational Potential to Employment	Location Variables	Number
Harmonic Centrality	Location Variables	Number
Closeness Centrality	Location Variables	Number
Betweenness Centrality	Location Variables	Number
Dwelling Type	Structural Variables	Categorical
Dwelling Size	Structural Variables	Number
Age of Building	Structural Variables	Number
Tenure	Structural Variables	Categorical
Number of Bedrooms	Structural Variables	Number
Newly Built	Structural Variables	Categorical
Number of Shops, Radius 800 Metres	Neighbourhood Variables	Number
Distance to Parks	Neighbourhood Variables	Number
Secondary School Score, Radius 800 Metres	Neighbourhood Variables	Number

A central concern in the cross-section OLS hedonic model was the unobserved time-invariant characteristic f in Equation 8, where unobserved factors were omitted. These factors included urban design, property aesthetics, composition of the neighbourhood effects and all of the factors that do not change significantly within the study's time frame. When the unobserved time-invariant characteristic was correlated with the independent variables, biases could occur in the estimates.

The panel data FE regression model was one method that could account for the unobserved variance in estimating the accessibility effects (Gibbons and Machin, 2005). This FE model examined the association between dependent and independent variables within the same property or postcode over time. By looking at within-differences, unobserved spatially associated time-invariant characteristics of the property were dropped. More simply, if an omitted variable did not change over time, then any change in Y could not be caused by the omitted time-invariant variable but rather by the time-variant characteristic. There are two main approaches to estimating the FE model. The foremost approach includes first differencing (FD: $X_{it} - X_{it-1}$) the variables in the standard regression model, where the unobserved effects were dropped ($f_i - f_i$). The second is the within-estimator, which was used for this study. The FE within-estimator uses a within-transformation, where both the independent and dependent variables are demeaned. The unobserved effects were also removed. The following is the general form of the within-transformation equation, following Gibbon's (2005) study. The first

¹⁸ An average postcode has 10-15 households. The sample average is 2.3 property transactions per household over a 15 year time period.

differencing estimator is equivalent to the FE within-estimator, where the time period (t) is equal to two.

$$(\text{Log}P_{it} - \text{Log}\bar{P}_i) = \beta_1(X_{it} - \bar{X}_i) + \beta_2(\text{Log}A_{it} - \text{Log}\bar{A}_i) + (g_t - \bar{g}) + (e_{it} - \bar{e}_i) + (f_i - \bar{f}_i)$$

Where

P is the dependent variable where i = postcode and t = year

P_bar is the average price

A represents the accessibility variable

A_bar is the average for each accessibility variable

X represents the vector of independent variables

X denotes averages for each independent variable

g_t is the general time trend variable

g is the average for the time trend variable

e_it is the error term

e is the averages of the error term

Equation 5.5

For this study, this equation was simplified further ($\tilde{X}_{it} = (X_{it} - \bar{X}_i)$).

$$\text{Log}(\tilde{P}_{it}) = \beta_1\tilde{X}_{it} + \beta_2\text{Log}(\tilde{A}_{it}) + \tilde{g}_t + \tilde{e}_{it}$$

Where

P-tilde is the demeaned of price at i = postcode and t = year

A-tilde represents the demeaned of the accessibility variable

X-tilde represents the demeaned of the independent variable

g-tilde is the demeaned of the time trend variable

e-tilde is the demeaned of the error term

Equation 5.6

The FE model allowed for a more robust estimation of the accessibility effects when compared to the cross-sectional OLS model (Gibbons and Overman, 2010; Gibbons and Machin, 2005). To test the effects of the transport projects, a FE model was estimated on the London house price dataset from 1995 to 2011. Both street and Tube network models had been constructed for the years 1995, 2000, 2005 and 2011. The transport projects for these four time periods included the London JLE, the London OG, the London DLR extension and the CRTL. Table 5.3 illustrates these transport projects from 1995 to 2011.

Table 5.3 London street-tube network model specifications from 1995 to 2011.

Spatial Network Model	London Street Network	London Underground and Railway, 1995	Jubilee Line Extension, 1999	Croydon Tramlink, 2000	London Overground East London Line, 2010	Docklands Light Railway Stage Two Extension, 1999	Docklands Light Railway Stage Three Extension, 2005
1995	Y	Y	N	N	N	N	N
2000	Y	Y	Y	Y	N	Y	N
2005	Y	Y	Y	Y	N	Y	Y
2011	Y	Y	Y	Y	Y	Y	Y

The regression models estimate 12 different accessibility measures identified in Table 1 by first using the pooled OLS regression model specified in Equation 5.4 and then using the FE regression model specified in Equation 5.6, controlling for all of the independent variables listed in Table 3.

Model 1 and Model 2 focused on estimating the gravitational potential of the hedonic price model. Model 3 and Model 4 concentrated on assessing the closeness centrality of the hedonic price model. Model 5 and Model 6 focused on assessing the harmonic centrality of the hedonic price model. Radius infinity and radius 60 minutes were calculated for all six models respectively. Table 5.4 shows the 12 OLS pooled regression models. Table 5.5 displays the 12 FE regression models. Table 5.6 exhibits the eight FE regression models where the accessibility measures were being tested jointly. The gravitational potential measure was used as the base measure (first with travel time impedance and then with angular impedance); the four geometric accessibility measures were added subsequently. Not all pair-wise accessibility measures were tested due to multicollinearity. A variation inflation factor (VIF)¹⁹, which tests for multicollinearity, was reported in the empirical section.

Table 5.4 OLS pooled regression models.

	Travel Time	Angular	Travel Time	Angular
	Radius Infinity		Radius 60 Minutes	
Gravitational Potential (GPi)	Model 1A	Model 2A	Model 1B	Model 2B
Harmonic Centrality (HCi)	Model 3A	Model 4A	Model 4B	Model 4B
Closeness Centrality (CCi)	Model 5A	Model 6A	Model 5B	Model 6B

Table 5.5 Fixed-effect regression models.

	Travel Time	Angular	Travel Time	Angular
	Radius Infinity		Radius 60 Minutes	
Gravitational Potential (GPi)	Model 1C	Model 2C	Model 1D	Model 2D
Harmonic Centrality (HCi)	Model 3C	Model 4C	Model 4D	Model 4D
Closeness Centrality (CCi)	Model 5C	Model 6C	Model 5D	Model 6D

¹⁹ Vif(i) = 1/1-r²

Table 5.6 Joint accessibility models.

	Travel Time	Angular	Travel Time	Angular
	Radius Infinity		Radius 60 Minutes	
Gravitational Potential (GPi)	Model All (E)			Model All (F)
Harmonic Centrality (HCi)	Model 3E	Model 4E	Model 3F	Model 4F
Closeness Centrality (CCi)	Model 5E	Model 6E	Model 5F	Model 6F

FE models and OLS models were typically estimated using an OLS estimator. Standard statistics, such as Bayesian information criterion (BIC)²⁰, were reported. The BIC is a goodness of fit statistic that compares different models and adjusts for the number of parameters. The BIC was calculated for all of the candidate models and compared; the lower the criterion, the better the model's quality.

5.4 Dataset and Study Area

5.4.1 Greater London Area

This section describes the dataset used for the empirical research. As described in Chapter 4, the analytical chapter used the Greater London area as the case study. Figure 5.4 displays the study area in black and the borough boundaries in red.



Figure 5.4 The area the Greater London case study.

²⁰ $BIC = \ln(n) k - 2\ln(LL)$

LL = maximum loglikelihood of model
k = number of parameters

5.4.2 House Prices Dataset

The house prices dataset, derived from Nationwide and the Land Registry²¹, was used for this analytical study; this dataset included a total of 130,484 transactions between 1995 and 2011. Figure 5.5 below describes house prices in London between 1995 and 2011, where the average house price rose more than four times over the last 15 years, from less than £80,000 to more than £350,000.

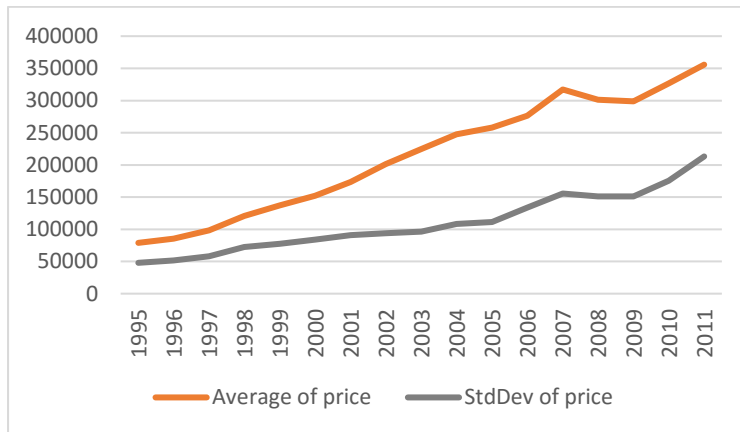


Figure 5.5 Average house prices in Greater London between 1995 and 2011.

Figure 5.6 below describes the overall average house price in Greater London for 1995, 2000, 2005 and 2010, visualised at the postcode level, where red denotes property price above £1,000,000 and blue denotes property price below £150,000.

²¹ The data was provided by the Nationwide through an agreement with London School of Economics. This is a subset of the open source Land Registry sold price dataset. The origins of all data on sold house prices in United Kingdom is owned by Land Registry/Registers of Scotland © Crown copyright 2013.

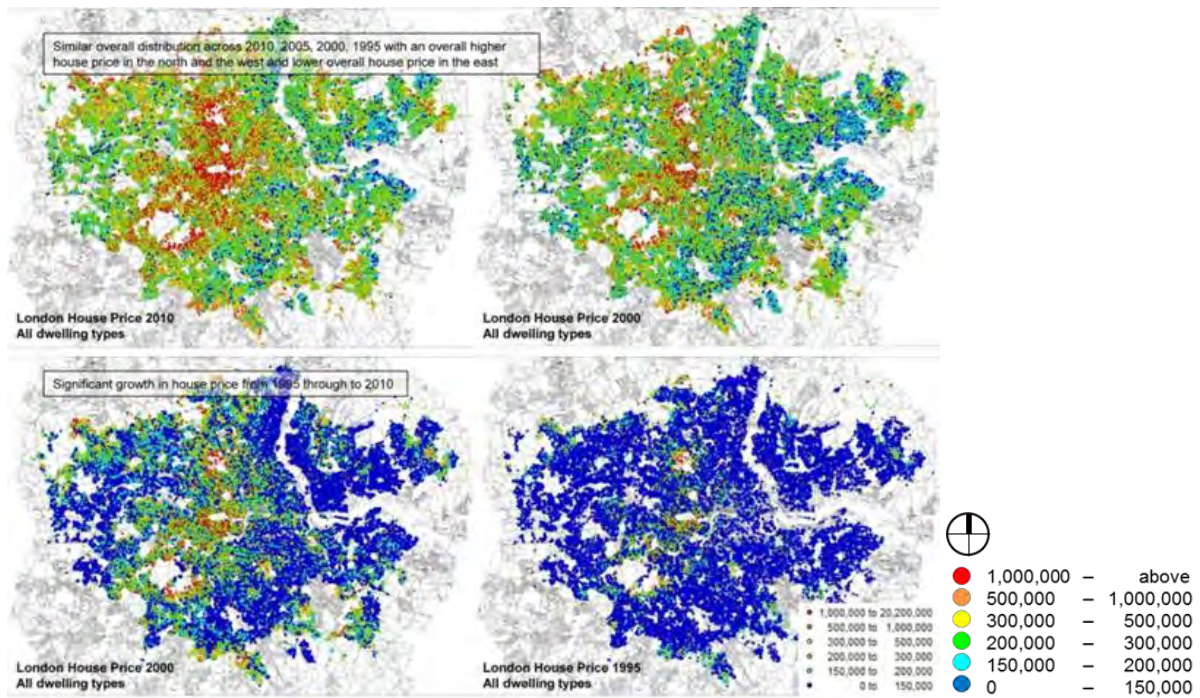


Figure 5.6 House prices in London in 2010, 2005, 2000, and 1995, visualised from red (high) to blue (low) using a constant colour range.

A key observation from the temporal distribution was the dramatic increase in house prices over time for the same location. The persistence of high house price areas contributed to a neighbourhood lock-in effect in the London housing market. Geographically, higher house prices were clustered near the centre of the city and in traditionally affluent areas such as Hampstead, Richmond, Kensington and Chelsea. Lower house prices were grouped in less central areas, such as Tottenham in the Northeast and Thamesmead in the Southeast. These house price distributions corresponded to previous research, which found that house prices relate to accessibility (Law et al., 2013; Ahlfeldt, 2010), attractive green space (Duncan, 2005), housing submarkets (Bourassa et al., 2008) and school quality (Gibbons and Machin, 2003).

Furthermore, Figure 5.7 compares the normalised house prices between 1995 (left) and 2011 (right), where red denotes relatively higher house prices and blue signifies relatively lower house prices.²² The maps for the two years are visualised in a grid, where the values are normalised and interpolated. The maps, therefore, show the relative increases rather than absolute increases. The result shows house prices in areas such as Southwark, Canary Wharf, Shoreditch, Hoxton and Hackney increased more than in other areas, which corresponded to a growth in accessibility. However, the study also showed areas such as the Royal Docks and Croydon had only moderate increases in house prices despite improvements in accessibility.

²² In order to compare the two years, house price are first normalised by $\text{norm}_p(it) = p(it) / \max_p(it)$ where norm_p is the normalised price, p is the price and $\max(p)$ is the maximum price.

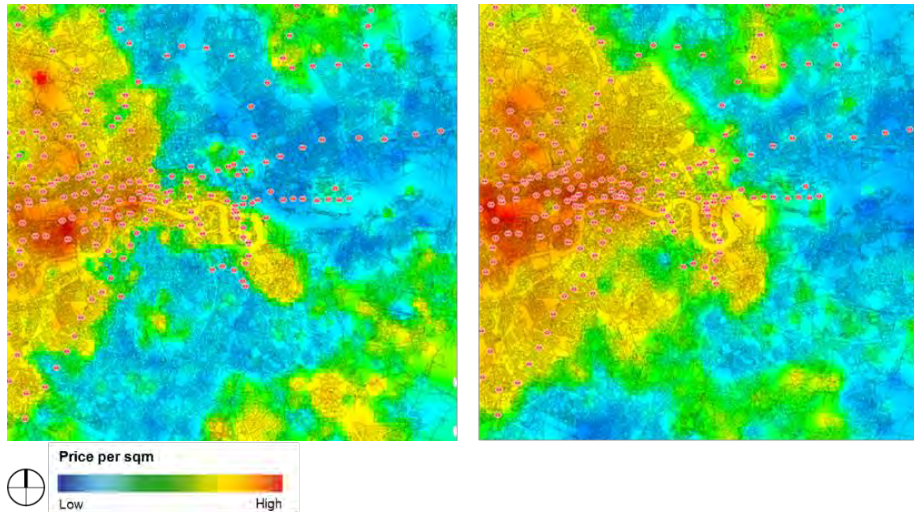


Figure 5.7 London house price per square metre in 1995 (left) and 2011 (right).

This result suggested different transport projects may have had different effects on house prices. To examine this further, Figure 5.8 shows house prices per square metre across time and within 800 metres of the four transport projects, namely the JLE in grey, the DLR in blue, the London OG in orange and the CTRL in green. The result showed the CTRL price per square metre increased significantly slower than the JLE, the OG and the DLR. These visual and descriptive results suggested that there were clear differences between transport projects in London. Transport links that improved accessibility to the city centre (e.g. JLE) seemingly had a higher house price effect than transport links that only improved outlying areas (e.g. CTRL).

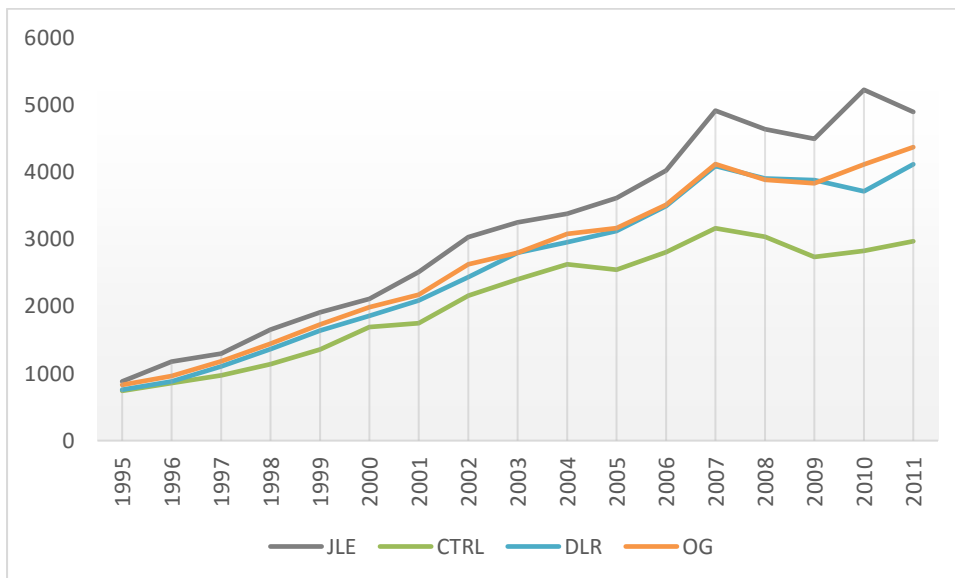


Figure 5.8 House prices per square metre comparison between the JLE, the CTRL, the DLR and the London OG.

5.4.3 Transport Innovation Projects Between 1995 and 2011

To capture the transport projects detailed in Figure 5.9, four street and Tube network models in London were constructed for the years 1995, 2000, 2005 and 2011.

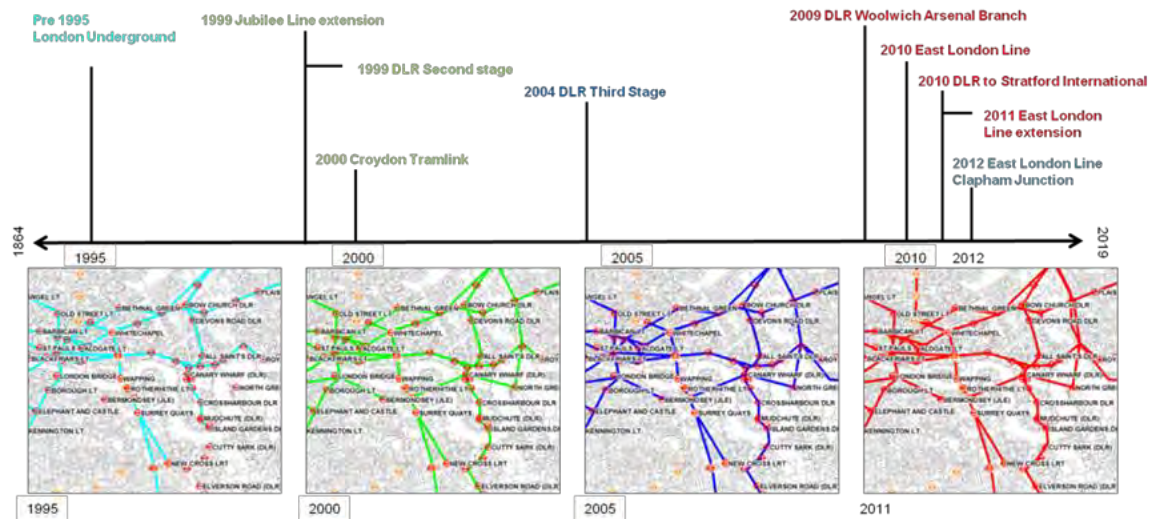


Figure 5.9 Transport projects in London between 1995 and 2011.

This study followed the Law et al. (2012) method of constructing the London Street and Tube network models, as illustrated in Chapter 4. The basis of the London street network was the Ordnance Survey Meridian 2 street network²³; the information for the London Tube network was retrieved from the Transport for London website (2016) and consisted of the London UG, the London OG, the DLR and the CRTL. The London Tube network was manually constructed by first modelling straight lines between all of the pairs of connected stations and then directly linking the result to the street model. There were 114,985 segments, of which 114,018 were street segments and 967 were the London Tube network segments. The normal street network had an assumed pedestrian speed of 5 km/h and the London Tube network had an assumed speed of 15 km/h. More information on this network can be found in Chapter 4. The three accessibility measures were calculated for the four spatial network models. Figure 5.10 shows the three accessibility measures, namely gravitational potential, closeness centrality and harmonic centrality, calculated for both the street and London Tube network models in 2011. The three measures were calculated for two radii and then split into angular cost (left) and travel time cost (right). The result showed the three measures were visually similar, with a clear concentric pattern where the accessibility core (spaces of highest accessibility in red) was in the centre of London (at the City of Westminster and the City of London). This result was expected, as the number of jobs has been shown to strongly correlate with the number of nodes in an area (Law et al. 2017). Despite the clear similarities, there were notable differences between the accessibility measures across the two radii and the two types of distances. The accessibility core was more

²³ Ordnance Survey Open Data Meridian 2 Dataset. © Crown Copyright {2014}

dispersed for radius infinity than it was for radius 60 minutes. The accessibility core was also more linear for angular impedance than for travel time impedance.

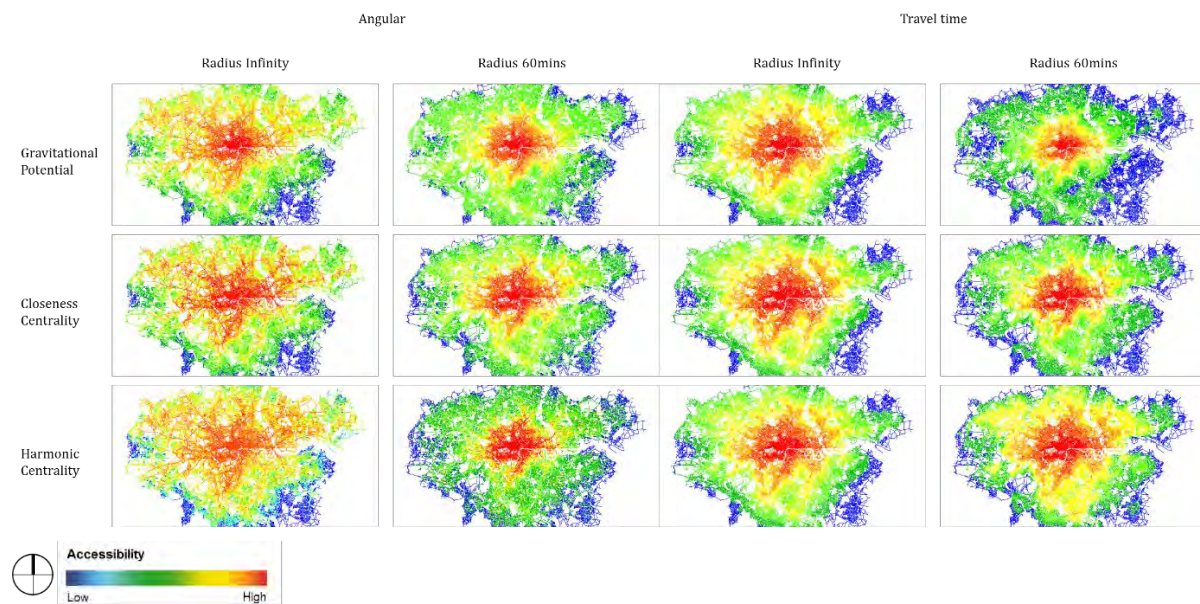


Figure 5.10 Accessibility measures mapped by a colour spectrum, where red denotes high accessibility and blue denotes low accessibility.

Figure 5.11 displays the closeness centrality measure in 1995 (left) and 2011 (middle) and a difference map²⁴ between the two (right). The difference maps for this research were calculated by taking the difference of the 2011 and 1995 closeness centrality values. Significant increases in centrality values were observed in East London between the two time periods. This was due to the inclusion of the JLE, the DLR and the London OG.



Figure 5.11 Closeness centrality measures for East London in 1995 (left), in 2011 (middle) and for the difference between the two years (right).

²⁴ $CC_diff = CC_2011 - CC_1995$

5.4.4 Descriptive Statistics

Table 5.7 describes the dataset used for this analytical chapter. This included sold house prices as the dependent variable, the accessibility measures as the response variables, and several structural and amenity factors as the controlled variables. The structural variables included size, age, the number of bedrooms and type of dwelling. The amenity variables included the number of shops within 800 metres²⁵, the metric network distance to a heritage park in London²⁶ and the average secondary school (A-level score) within 800 metres.²⁷ The calculation of these controlled variables is described in Chapter 4.

Table 5.7 Descriptive statistics for the variables used in the empirical analysis.

Variables	Descriptions	-1 N	-2 Mean	-3 Standard Deviation	-4 Minimum	-5 Maximum
Price	Sold Price	130,484	193,167	137,964	8,000	4.40E+06
Bedrooms	Number of Bedrooms	130,484	2.428	0.936	1	10
Floors	Floor Size	130,484	92.02	36.02	24	278
Age	Age of Dwelling	130,484	74.5	35.22	0	497
ctype_dum1	Converted Flat	130,484	0.194	0.395	0	1
ctype_dum2	Converted Maisonette	130,484	0.0151	0.122	0	1
ctype_dum3	Cottage	130,484	3.83E-05	0.00619	0	1
ctype_dum4	Detached	130,484	0.0377	0.191	0	1
ctype_dum5	Detached Bungalow	130,484	0.00465	0.068	0	1
ctype_dum6	Purpose Built Flat	130,484	0.212	0.409	0	1
ctype_dum7	Purpose Built Maisonette	130,484	0.00795	0.0888	0	1
ctype_dum8	Semi-bungalow	130,484	0.00677	0.082	0	1
ctype_dum9	Semi-detached	130,484	0.211	0.408	0	1
ctype_dum10	Terraced	130,484	0.311	0.463	0	1
new_build_dum1	Not New Build	130,484	0.973	0.162	0	1
new_build_dum2	New Build	130,484	0.0268	0.162	0	1
tenure_dum1	Freehold	130,484	0.565	0.496	0	1
tenure_dum2	Leasehold	130,484	0.435	0.496	0	1
avg_alevel_800	Average A-level score within 800m	130,484	366	389.1	0	1,349
met_park	Metric Network Distance to Park in London	130,484	2,376	1,801	0	9,809
active_r800	Number of Shops within 800m	130,484	98.81	100.3	0	1,204
global_ch	Global Angular Betweenness	130,484	1.76E+07	6.88E+07	0	1.66E+09
grav_ang_rn	Angular Gravitational Potential Radius N	130,484	279,826	57,823	112,402	615,835
grav_time_rn	Travel Time Gravitational Potential Radius N	130,484	38,645	14,779	11,178	142,608
grav_ang_r60	Angular Gravitational Potential Radius 60 minutes	130,484	47,550	71,202	47.79	473,671
grav_time_r60	Travel Time Gravitational Potential Radius 60 minutes	130,484	9,919	15,610	5.774	134,026
cc_ang_rn	Angular closeness centrality Radius N	130,484	7,319	1,140	3,425	10,849
cc_time_rn	Travel Time Closeness Centrality Radius N	130,484	782.4	174.6	322.2	1,268

²⁵ Shops are classified under the retail category in the Valuation Office Agency's business rates data. Data provided by the Valuation Office Agency contains public sector information licensed under the Open Government Licence v1.0.

²⁶ This includes all parks in London

²⁷ The A-Level (General Certificate of Education Advanced Level) is an academic qualification offered by educational institutions in England, Wales and Northern Ireland to students completing secondary or pre-university education.

cc_ang_r60	Angular Closeness Centrality Radius 60 minutes	130,484	997.6	777.8	13.53	4,723
cc_time_r60	Travel Time Closeness Centrality Radius 60 minutes	130,484	170.8	148.3	1.61	850.3
hc_ang_rn	Angular Harmonic Centrality Radius N	130,484	74,250	7.86E+06	3,623	1.00E+09
hc_time_rn	Travel Time Harmonic Centrality Radius N	130,484	981.7	257.5	343.2	1,891
hc_ang_r60	Angular Harmonic Centrality Radius 60 minutes	130,484	67,222	7.86E+06	17.35	1.00E+09
hc_time_r60	Travel Time Harmonic Centrality Radius 60 minutes	130,484	205	167.4	3.052	1,083
year	Year of Transaction	130,484			1995	2011
postcode	Postcode Number	56,873				

5.5 Empirical Results

5.5.1 The OLS Regression Model Results

This research first applied the pooled OLS regression model in Equation 5.4 to compare the six accessibility measures. Table 5.8 shows the results of the accessibility measures at radius infinity, and Table 5.9 shows the results for the accessibility measures at radius 60 minutes. Estimates of the controlled variables were not reported, as they conformed to existing literature where variables such as the floor size, the dwelling type and age, the number of shops, the school quality and the distances to parks were all significant variables with the expected signs. The pooled OLS model achieved an overall fit where approximately 80% of the variation in the house prices could be explained by the model. The two sets of models achieved similar goodness of fits, with the radius 60 minutes measure attaining a higher r^2 and a lower BIC . This result showed accessibility measures at radius 60 minutes were preferred to the same measure at radius infinity. In the model where the P-value ($\text{Prob} > f$) was less than 0.01, F-tests showed significance in all of the accessibility variables. The significance of the geometric and geographic accessibility variables showed positive effects on house prices.

Gravitational potential to employment achieved a lower BIC than the two geometric accessibility measures and was preferred to using the OLS model. The estimates for the two sets of OLS models were approximately 0.3–0.4. The research also showed that there were small differences between the closeness centrality and harmonic centrality measures. Not shown in the tables, the betweenness centrality measure displayed a negative relationship with house prices in all of the models. In the next section, the research adopts the FE regression model to test the extent to which the public transport projects' accessibility improvements influenced house prices.

Table 5.8 Pooled regression model for radius infinity.

Variables	1 Model 1b	2 Model 2b	3 Model 3b	4 Model 4b	5 Model 5b	6 Model 6b
Control var. Year (trends)	Yes	Yes	Yes	Yes	Yes	Yes
GravAngRN	0.448*** (0.0045)					
GravTimeRN		0.368*** (0.00283)				
CCAngRN			0.376*** (0.00574)			
CCTimeRN				0.390*** (0.00442)		
HCAngRN					0.137*** (0.00366)	
HCTimeRN						0.378*** (0.00391)
Constant	4.843*** (-0.0563)	6.597*** (-0.0301)	7.082*** (-0.0511)	7.843*** (-0.0298)	9.171*** (-0.0335)	7.844*** (-0.0273)
BIC	35,029	28,670	40,341	36,977	43,171	35,500
Observations	130,484	130,484	130,484	130,484	130,484	130,484
R-squared	0.821	0.829	0.814	0.818	0.809	0.82

*** p<0.01, ** p<0.05, *p<0.

Table 5.9 Pooled regression model for radius 60 minutes.

Variables	-1 Model 1a	-2 Model 2a	-3 Model 3a	-4 Model 4a	-5 Model 5a	-6 Model 6a
Control var. Year (trends)	Yes	Yes	Yes	Yes	Yes	Yes
GravAngR60	0.112*** (0.00082)					
GravTimeR60		0.113*** (0.00082)				
CCAngR60			0.126*** (0.00147)			
CCTimeR60				0.129*** (0.00137)		
HCAngR60					0.0760*** (0.00129)	
HCTimeR60						0.139*** (0.00151)
Constant	9.349*** (0.0105)	9.522*** (0.00951)	9.588*** (0.012)	9.809*** (0.00965)	9.891*** (0.0113)	9.733*** (0.0103)
BIC	27,220	26,655	37,403	35,993	41,113	36,341
Observations	130,484	130,484	130,484	130,484	130,484	130,484
R-squared	0.831	0.832	0.818	0.82	0.812	0.819

*** p<0.01, ** p<0.05, * p<0

Exploring this model further, a pooled OLS regression model was applied to the area within 800 metres of the following transport projects: the JLE, the DLR and the CTRL. The London OG was not included in this analysis, as it was not built in a similar time period. Table 5.10 shows the OLS results using the angular closeness centrality measure with the controlled variables. The results showed that accessibility, which in this case was the closeness centrality, had a significantly higher estimate for the JLE than both the DLR and the CTRL. This result suggested various areas and transport systems in London had different economic values for accessibility.

Table 5.10 Pooled regression model for the three transport lines in London.

Variables	-1 Jubilee Line Extension	-2 Croydon Tramlink	-3 Docklands Light Railway
Control var.			
Year (trend)	Yes	Yes	Yes
CCAngR60	0.380*** (0.0326)	0.113*** (0.0116)	0.176*** (0.0114)
Constant	7.553*** (0.254)	9.807*** (0.0805)	8.958*** (0.0948)
BIC	1,198	-446	1,513
Observations	1,564	4,571	3,002
R-squared	0.785	0.873	0.815

*** p<0.01, ** p<0.05, * p<0

5.5.2 Fixed-Effect Regression Results

This section applied the FE regression model based on Equation 5.6 to test the extent to which accessibility improvements from the transport projects influenced house prices in London.²⁸ The FE regression model achieved an overall fit where the regression model could explain approximately 90% of the variation in the house prices. The result showed a marked improvement from the pooled OLS model. Due to the presence of heteroskedasticity²⁹, the standard errors were estimated using a robust least squares estimator, following standard procedures (Osland and Thorsen 2008). Table 5.11 summarises the first six candidate models, which were based on RN. The results showed the signs were not in the expected direction, where higher house prices were found in less accessible space. Under the more robust and conservative specifications, the findings showed that accessibility measures at radius infinity did not influence house prices.

Table 5.11 Fixed-effect regression results for radius infinity.

Variables	-1 Model 1b	-2 Model 2b	-3 Model 3b	-4 Model 4b	-5 Model 5b	-6 Model 6b
Control var.	Yes	Yes	Yes	Yes	Yes	Yes
Year(trend)						
GravAngRN	-0.262*** (0.0609)					
GravTimeRN		-0.197*** (0.0257)				
CCAngRN			-0.530*** (0.0666)			
CCTimeRN				-0.433*** (0.0337)		
HCAngRN					-0.467*** (0.0801)	
HCTimeRN						-0.222*** (0.0312)
Constant	14.03*** (0.759)	12.81*** (0.269)	15.45*** (0.591)	13.63*** (0.224)	14.96*** (0.721)	12.28*** (0.214)
BIC	-204,338	-204,420	-204,466	-204,639	-204,380	-204,395
Observations	130,484	130,484	130,484	130,484	130,484	130,484
R-squared	0.917	0.917	0.917	0.917	0.917	0.917
Postcode (n)	56,873	56,873	56,873	56,873	56,873	56,873

*** p<0.01, ** p<0.05, * p<0.1

²⁸ The fixed effect model is used rather than the random effect model as the null hypothesis for the Hausman test was rejected. An important assumption of the FE model is that time-invariant characteristics are unique to the individual and should not be correlated with other individual characteristics. In empirical research, a Hausman test is often applied for model selection.

²⁹ Wald statistic for group-wise heteroskedasticity shows the hypothesis test for homogeneity have been rejected suggesting heteroskedasticity exist.

Table 5.12 summarises the next six FE regression models, with accessibility measures at radius 60 minutes. The f-test showed significance on the six accessibility measures and all of the signs pointed in the expected direction. The results showed the accessibility improvements from the public transport projects had a positive effect on the house prices in London. This result differed significantly from the radius infinity measures, where the signs were not in the expected direction. Like the OLS model, the findings showed that the accessibility measures at radius 60 minutes were preferred to the same measures at radius infinity. Under this more conservative and robust method, the estimates for closeness centrality dropped to 0.16 and gravitational accessibility estimates fell to 0.05. This result showed that the specifications of the radius and travel time budget clearly influenced the results. The differences in the estimates also suggested that the two types of accessibility may have had different effects on house prices. The findings also showed that there were minor differences when calculating accessibility using either angular distance or travel time distance.

Table 5.12 Fixed-effect regression results for radius 60 minutes.

Variables	1 Model 1a	2 Model 2a	3 Model 3a	4 Model 4a	5 Model 5a	6 Model 6a
Control var.	Yes	Yes	Yes	Yes	Yes	Yes
Year (trend)						
GravAngR60	0.0532*** (0.00681)					
GravTimeR60		0.0506*** (0.00633)				
CCAngR60			0.156*** (0.0165)			
CCTimeR60				0.111*** (0.0125)		
HCAngR60					0.166*** (0.0177)	
HCTimeR60						0.126*** (0.014)
Constant	10.24*** (0.0698)	10.35*** (0.056)	9.724*** (0.112)	10.22*** (0.0649)	9.619*** (0.124)	10.12*** (0.0748)
BIC	-204,452	-204,454	-204,513	-204,480	-204,510	-204,483
Observations	130,484	130,484	130,484	130,484	130,484	130,484
R-squared	0.917	0.917	0.917	0.917	0.917	0.917
Postcode(n)	56,873	56,873	56,873	56,873	56,873	56,873

*** p<0.01, **p<0.05, *p<0.1

5.5.3 Fixed-Effect Regression Model Results for Testing Joint Accessibility Effects

To examine the joint effect, this research estimated eight joint models, which included both geographic and geometric accessibility measures. Only radius 60 minutes was tested due to the significance in both the OLS and FE models. Table 5.13 summarises the results for the models. Models 1 to 4 used gravitational potential at radius 60 minutes, with travel time impedance as a base model. In all four models, both the geographic and the geometric accessibility variables were significant. The estimates for closeness centrality and harmonic centrality were 0.12, while for the gravitational potential, the estimate was 0.02. The estimations represented the lower bounds and were expectedly lower than the OLS specification. Models 5 to 8 used gravitational potential at radius 60 minutes, with angular impedance as a base model. In all four models, both the geographic and geometric accessibility variables were similarly significant with comparable estimates. These results

showed that the transport projects had substantial cumulative effects on house prices in London between these time periods. Furthermore, these results showed that the geographic and geometric accessibility measures largely had overlapping effects on house prices, which was seen in the reduced significance of the estimates for both measures; each had a similar goodness of fit. This result showed geometric accessibility measures were useful proxies for the geographic accessibility measures in the hedonic price models, where the two measures explained similar variations in house prices. The result also showed a weak but significant differential effect between the two accessibility measures; this was demonstrated by the joint significance when the gravitational potential measures and the geometric accessibility measures were included in the hedonic price model.

Table 5.13 Joint accessibility fixed-effect regression model results.

Variables	1 Model 1c	2 Model 2c	3 Model 3c	4 Model 4c	5 Model 5c	6 Model 6c	7 Model 7c	8 Model 8c
Control var.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year (trend)								
GravAngR60	0.0192** (0.0088)	0.0258*** (0.0088)	0.0200** (0.0088)	0.0259*** (0.0086)				
GravTimeR60					0.0177** (0.0084)	0.0239*** (0.0085)	0.0185** (0.0083)	0.0238*** (0.0083)
CCAngR60	0.123*** (0.0211)				0.123*** (0.0217)			
CCTimeR60		0.0764*** (0.0159)				0.0759*** (0.0166)		
HCAngR60			0.129*** (0.0225)				0.129*** (0.0231)	
HCTimeR60				0.0883*** (0.0174)				0.0877*** (0.0182)
Constant	9.758*** (0.112)	10.14*** (0.0745)	9.678*** (0.125)	10.06*** (0.0801)	9.800*** (0.115)	10.20*** (0.0666)	9.723*** (0.129)	10.12*** (0.0749)
BIC	-204,513	-204,488	-204,510	-204,493	-204,512	-204,486	-204,509	-204,490
Observations	130,484	130,484	130,484	130,484	130,484	130,484	130,484	130,484
R-squared	0.917	0.917	0.917	0.917	0.917	0.917	0.917	0.917
Postcode (n)	56,873	56,873	56,873	56,873	56,873	56,873	56,873	56,873

*** p<0.01, **p<0.05, *p<0.1

5.5.4 Multi-collinearity tests between geometric and geographic accessibility measures

To ensure the accessibility measures did not violate multicollinearity, a VIF test on the accessibility measures was employed. As a rule-of-thumb, parameters could be included if $VIF < 10$ and excluded if $VIF > 10$. Table 5.14 shows the VIF matrix for all pair-wise combinations of the six accessibility measures. The results clearly showed that the same accessibility measures, with both angular and travel time impedance, should not be jointly included; closeness centrality and harmonic centrality should also not be jointly included. The other combinations could be included under this VIF threshold. Despite being below the threshold, the results clearly demonstrated that geometric and geographic accessibility measures strongly overlap and correlate with one another. Dimension reduction techniques such as Principal Component Analysis (PCA) can be use in the future to minimise collinearity biases.

Table 5.14 Variation inflation factor (VIF) measured for the six accessibility measures. VIF>10 is highlighted in red.

VIF_matrix	GravAngR60	CCTimeR60	CCAngR60	HCTimeR60	HCAngR60
GravTimeR60	56.25	9.83	7.12	9.85	2.83
GravAngR60		9.44	9.17	9.23	3.21
CCTimeR60			30.62	193.27	3.73
CCAngR60				26.79	4.75
HCTimeR60					3.75

5.6 Discussion

To summarise, the results show that large-scale transport projects built in London between 1995 and 2011 have had significant effects on house prices. Empirically, using the OLS specification and the more robust FE specification, both the geographic and geometric accessibility variables are significant when associated with house prices in London. The results extend from the established relationship between accessibility measures and house prices by comparing geographic and geometric accessibility measures, two types of distance and two types of radii. This section provides a general discussion and plausible interpretation of the research findings.

The first key finding of the research shows robust evidence using a panel data regression model confirming geographic and geometric accessibility effects on house prices from key transport projects in London between 1995 and 2011. This is to be expected, as the increase in accessibility reduces travel time and increases the access to job opportunities, amenities, space and people.

The second key finding of the research shows a strong overlap across geographic and geometric accessibility measures. A possible explanation is illustrated in Hillier's (1996) theory of the movement economy, where the grid is a generator of movement and land use. Under this analytical approach, spatial configuration measures predict employment location, where accessibility-driven land use acquires the most accessible space. This overlap allows geometric accessibility to be used, especially when data for geographic measures are not available. This is particularly useful in the UK, where the official Census data is only available in aggregated format.

The third key finding of the research reveals the two accessibility measures have subtle differential effects on house prices. There are several plausible reasons why the two measures may differ in explaining house prices. First, when purchasing a property, buyers do not only purchase the access to employment, shops, parks and schools, but also to spatial network attractions, such as greater connectivity. This positive street network effect can come from the greater social interaction opportunities afforded by the denser grid in the centre (Jacob, 1961) or the generic accessibility effects that are not captured by geographic accessibility measures (Webster, 2010). These differences are, however, minor as there is a strong overlap between the two accessibility measures. Places with the greatest connectivity and opportunity for interaction are also strongly associated with the number of employment opportunities (Law et al 2015). An alternative interpretation for the

differential effect is that geometric accessibility may be capturing the latent potential of the location. This is logical as the formation of a desirable neighbourhood and the supply of residential space often takes longer than a new piece of transport infrastructure. Both interpretations require further research for validation.

The fourth key finding of the research shows that house prices correlate positively with spatial network closeness centrality but negatively with spatial network betweenness centrality. This is logical, as buyers are not simply purchasing accessibility to central places, but also quieter spaces that are protected from the noise and pollution of high betweenness spaces. Qualitatively, this effect was stronger along Green Lanes and weaker at Crouch Hill, suggesting subtle differences in the urban environment.

The fifth key finding of the research is that the accessibility measures for radius 60 minutes are preferred to the accessibility measures for radius infinity. This result is logical, as, in South East England, the average commuting time is between 50 to 70 minutes. This research shows the importance in specifying a radius cut-off for measuring accessibility and where more research is required to test different radii.

The last finding of the research is that the results of the angular and travel time impedances are not consistent. For the OLS model, travel time impedance appears to achieve a lower BIC, whereas, for the FE model, angular impedance seems to have a slightly lower BIC using geometric accessibility measures. Thus, the results are inconclusive and, again, further research is required.

5.6.1 Limitations

Despite the novelty of applying both geographic and geometric accessibility measures in hedonic price models, there are several limitations. The descriptive results showed there are key differences across the public transport projects in London, where there appear to be greater increases in house prices near the JLE compared to the CRTL. Thus, there is a need to identify the extent to which these transport projects differently influenced house prices and how various types of transport projects, such as the Cycle Superhighway or new bus links, can influence house prices in London. Future research is also needed to encompass inter-regional transport innovation projects. This is important as future regional transport projects such as the HS2 (High Speed 2 Railway) project will make inter-city commuting feasible, whereby the housing markets between London and Birmingham may overlap and influence each other.

Second, there is a need to better understand the actual processes through which accessibility effects influence house prices. Future research needs to observe and focus on individual residential location choices to examine the process of geometric accessibility influencing house prices. This can be accomplished through a residential location choice survey and analysis.

Third, accessibility improvements may not affect house prices homogeneously, as improvements may not necessarily translate to actual commuting behaviour changes. For example, in a car-dependent

neighbourhood, accessibility improvements might not induce modal change and therefore will increase house prices. As a result, further research that looks at commuting behaviour rather than potential accessibility should be investigated. This research also focused on the public transport accessibility effects in London. A more comprehensive study that studies the extent private transport accessibility and public transport accessibility effects differ in London is necessary.

Fourth, as the research only focused on the Greater London Area, further research across geographical regions, periods of time and different submarkets is needed to ensure the results can be generalised and better understood. This evidence can begin to reveal differences in the value of accessibility across time, groups and regions. For example, young professionals and students may find it more attractive to live near the centre, while families may find it more enticing to live near a good school or in a house with a large garden. Hence, research is required to examine the accessibility effects for different cities, demographic groups, and morphological structures across various time periods.

5.6.2 Conclusion

To end, the analytical results show accessibility has a significant effect on house prices in London. The use of geometric accessibility measures provides not only an objective method to account for accessibility but also adds a new dimension to understanding the spatial network accessibility effects on house prices. This research finds geometric accessibility effects overlaps significantly with geographic accessibility measures, such as gravitational potential. The study also finds significant but minor differential effects on house prices between the geographic and geometric accessibility measures. Specifying both accessibility measures in the hedonic price model allows for more informed policies on how changes in spatial configuration and employment locations associate with house prices. Further research is needed across geographical regions, modal split and time periods to further disentangle these effects (Law et al. 2017a). This research direction is briefly discussed in the last chapter. The next analytical chapter explores the Street-based Local Area effect on house prices.

Chapter 6

Street-based Local Area Effects on House Prices

6.1 Introduction

Research examining intra-city house price variations, using the hedonic price approach, often focused on estimating the implicit price at which buyers and sellers were willing to exchange contracts for elements like structural features, accessibility levels and neighbourhood amenities (Rosen, 1974; Cheshire and Sheppard, 1998). Applying the hedonic price approach, the previous chapter demonstrated that both geographic and geometric accessibility variables were found to be significant when associated with house prices in London between 1995 and 2011. The results confirmed the established relationship between geometric accessibility measures and property values (Law et al., 2013; Desyllas, 1997; Chiaradia et al., 2012; Yang, et al., 2015). However, this thesis argues that location differentials in house prices are captured not only by accessibility effects but also by local area effects, as defined by the street network. This conclusion follows previous spatial configuration research, which suggested that the topology of a street network not only relates to how we move in space but also how we associate with a place (Dalton et al., 2006; Yang and Hillier, 2007). This research proposes the concept of street-based local area (St-LA) and aims to test the extent to which St-LA has an effect on house prices. Using the 2011 London house prices dataset, this study employs a multilevel hedonic price approach to estimate how the St-LA affects house prices. The remainder of the chapter is organised as follows. Section 1 introduces previous research on local area units. Section 2 introduces the framework for defining St-LA; Section 3 provides detail about how the multilevel, empirical hedonic price method is applied to answer the research question. Section 4 introduces the London case study and the hedonic price model dataset. Section 5 reports the estimated results, and Section 6 discusses the findings and limitations of the study.

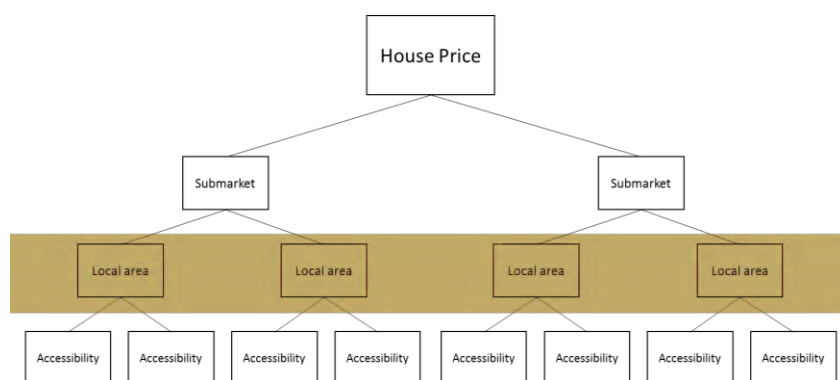


Figure 6.1 This chapter focuses on the St-LA effect on house prices.

6.1.1 Strand Two Research Question Definition

An under-explored topic within the field of urban planning and housing studies is that of the local area unit. A local area unit, here, is defined as a geographical unit that is larger than the immediate home

area (property boundary) but smaller than the city (Kearns and Parkinson, 2001). This unit is related to the concept of 'neighbourhood' in urban studies, which encompasses more complex historical, socio-economic and perceptual constructs that overlap according to the geographical scale (Lebel et al., 2007; Galster 2001; Kearns and Parkinson, 2001).

Census tracts or ward boundaries are administrative region-based local area units that are commonly used to capture neighbourhood characteristics. Due to their convenience, these boundaries have often been used in estimating hedonic price models or in defining housing sub-markets (Orford 2000; Goodman and Thibodeau, 1998; 2002). However, these local area units are seen as arbitrary, as they cut across streets and buildings; researchers recognise that these definitions do not necessarily capture the physical boundaries of a neighbourhood (Coulton et al., 2001). Figure 6.2 shows an area in London known as the Isle of Dogs overlaid with a Middle Layer Super Output Area (MSOA) taken from the UK census boundary. The map shows that the boundaries of the MSOA (red) cut across Canary Wharf's central office areas (blue).

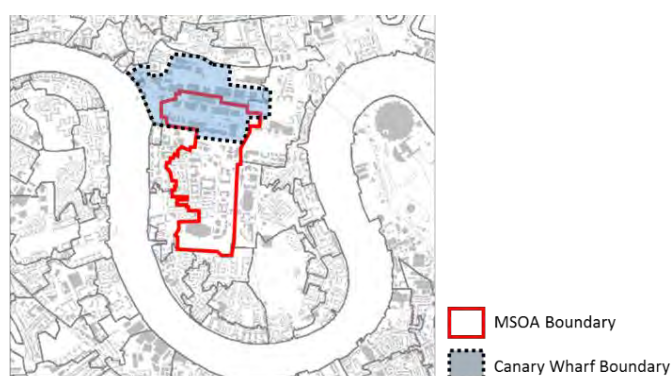


Figure 6.2 The Canary Wharf boundary (blue) overlaid with the MSOA boundary (red).

One problem of these 'arbitrary' or 'ad-hoc' (Orford, 1999; Goodman, 1977) administrative local area units is that they create inconsistent empirical results. Using New Haven as a case study, Goodman's studies (1978; 1981) found differences in coefficient estimates when comparing block level and census tract level. In 1985, this time using Baltimore as a case study, Goodman found segregation indices differed when applied across various aggregation levels. Differences between census tracts and the smaller block-level aggregation have been attributed to the 'fuzziness' and 'arbitrariness' of these local geographies. These problems extend to the identification of housing submarkets, as noted by Leishman (2009). For example, Bourassa et al. (1999) compared housing submarkets in both Sydney and Melbourne, which were defined using either data on individual dwellings or census tract-level data. This research found that different results were generated by grouping dwelling data than by grouping census tract data. These early studies found inconsistencies when calculating segregation indices, when estimating hedonic price regression models and when defining housing submarkets. Recent research also suggests resident perception maps of neighbourhoods could be more meaningful than administrative boundaries (Coulton et al. 2001). It is for these reasons that this research will propose the concept of an St-LA via asking the following research question:

To what extent do St-LAs, as defined by the topology of the street network, associate with house prices? Secondly, how do St-LAs compare with administrative local area units when correlating with house prices?

6.2 Framework for Street-based Local Area

A street-based local area is defined as a local area that is: 1) street-based, 2) topological or configurational, 3) has membership in discrete form, and 4) is larger than a home area but smaller than a city. The concept of St-LA stems from two fields. The first is network science, from which it borrows the concept of community structure, which is a characteristic found in many social and biological networks (Girvan and Newman 2002). Second is space syntax research, from which it derives the use of a spatial network dual graph in representing a city.

This research suggests that St-LAs significantly associate with house prices and that the St-LAs achieve a greater goodness of fit than traditional region-based local. There are two possible reasons for this. First, residents perceive the local area as distinct from the street network. Therefore, the street network is able to capture subtle differences and definitions of the urban environment more precisely than ad-hoc administrative regions. Second, the topology of the street network reinforces socio-economic similarities over time. As people identify with these local areas, this would have an effect on house prices. To study the research question, this chapter will compare how St-LAs and traditional administrative region-based local areas associate with house prices. Figure 6.3 compares each area type for the Isle of Dogs.

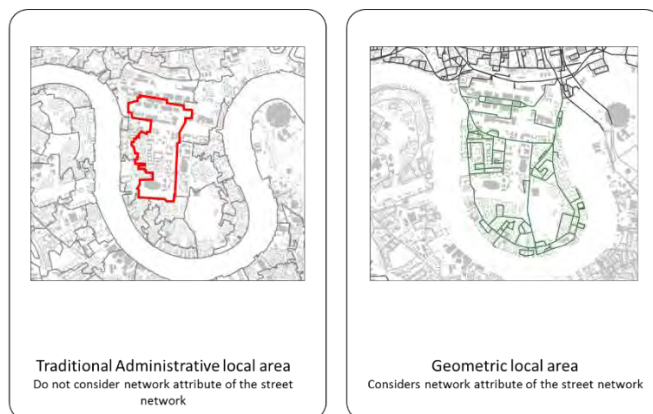


Figure 6.3 Traditional administrative local area (left) and St-LA (right).

The next section will describe the community detection techniques used to identify the St-LAs for this study.

6.2.2 Community Detection Methods

The objective of community detection is to define a set of subgraphs that maximise internal ties and minimise external ties by strictly using the graphs' topologies. These techniques found strong associations with social, functional and geographical network groupings (Girvan and Newman, 2002;

Guimer'a et al., 2005; Caschili et al., 2009). A key reason for the use of community detection techniques in defining St-LAs is that spatial network clustering can plausibly be related to socio-economic clustering or to perceptual homogeneity found in neighbourhoods or local areas. Previous research did not apply such techniques on the street network to find locality. Therefore, this research is novel in its application of community detection techniques on the street network dual graph.

6.2.3 Defining St-LA Using the Street-network Dual Graph

In graph theory, a street network is a type of planar graph embedded in Euclidean space. Two types of spatial network graphs can be identified, including the spatial primal graph, whose vertices are junctions and edges are streets, and the spatial dual graph **DG**, whose vertices **u** are streets and edges **e** are junctions (Porta et al., 2006). The latter has been popularised by space syntax research (Hillier and Hanson, 1984).

$$DG(u, e)$$

where

u is the node (street segments)

e is the edge (junctions)

Equation 6.1

This study will employ the community detection technique on the spatial dual graph of the road centre line in defining St-LA (Turner, 2007). More formally, an St-LA is defined as a discrete subgraph (subset) of the spatial DG, where all vertices (streets) classified within each subgraph share a membership.

$$SG_k \subseteq DG \text{ where } k = 1, 2, \dots, K$$

where

SG is the subgraph

DG is the spatial dual graph

K is the number of subgraph

Equation 6.2

One justification for the use of the dual graph representation is that a property is located on a street rather than on a junction. Community detection on a primal graph identifies clusters of connected junctions rather than groups of connected streets. The next section will describe the community detection method that identifies the subgraph.

6.2.4 Modularity Optimisation Algorithm on the Street-network Dual Graph

A large amount of research has been conducted into identifying community structures. Many algorithms have been proposed, including the modularity-based algorithm, the spin-glass algorithm, the Walktrap algorithm, the betweenness cut algorithm and the vertex propagation algorithm (Reichart and Bornholdt, 2004; Raghavan, et al., 2007; Newman and Girvan 2004; Pons and Latapy, 2006). These algorithms were briefly introduced in Chapter 3. This study, in particular, adopts the multi-level modularity optimisation algorithm and applies it to the street-network dual graph to identify the St-LAs.

This algorithm was tested in Chapter 3 and is one of the most commonly used in network science, as it is known for its efficiency and accuracy (Blondel et al., 2008; Lancichinetti and Fortunato, 2009).

This algorithm optimises against a community quality function called Modularity, which calculates the difference between the observed number of edges and the expected number of edges within a subgraph. The higher the number of observed edges relative to the number expected, the greater the modularity. As a result, this algorithm finds an optimal spatial network grouping that has more internal connections and fewer external connections. See Chapter 3 for more details.

6.2.5 Street-based Local Area Subgraph Network

In order to assess inter-cluster house price variations, an inter-cluster St-LA subgraph network was constructed based on the St-LA. The St-LA subgraph network is more formally defined as $G_{sub}(u, e)$, where nodes u are the St-LAs and e are the connections between all St-LAs.

$$G_{sub}(u, e)$$

where

U are the St-LA

E are the connections between St-LA

Equation 6.3

Figure 6.4 below illustrates the construction of the street-based subgraph network. From left to right, the first image shows the original spatial network graph. The second displays the St-LA membership that was found by using the method described in section 2.1. The third image presents the subgraph centroid, and the fourth image exhibits the edges between all of the connected subgraphs, where the width of the edges denote the connectivity between the subgraphs.

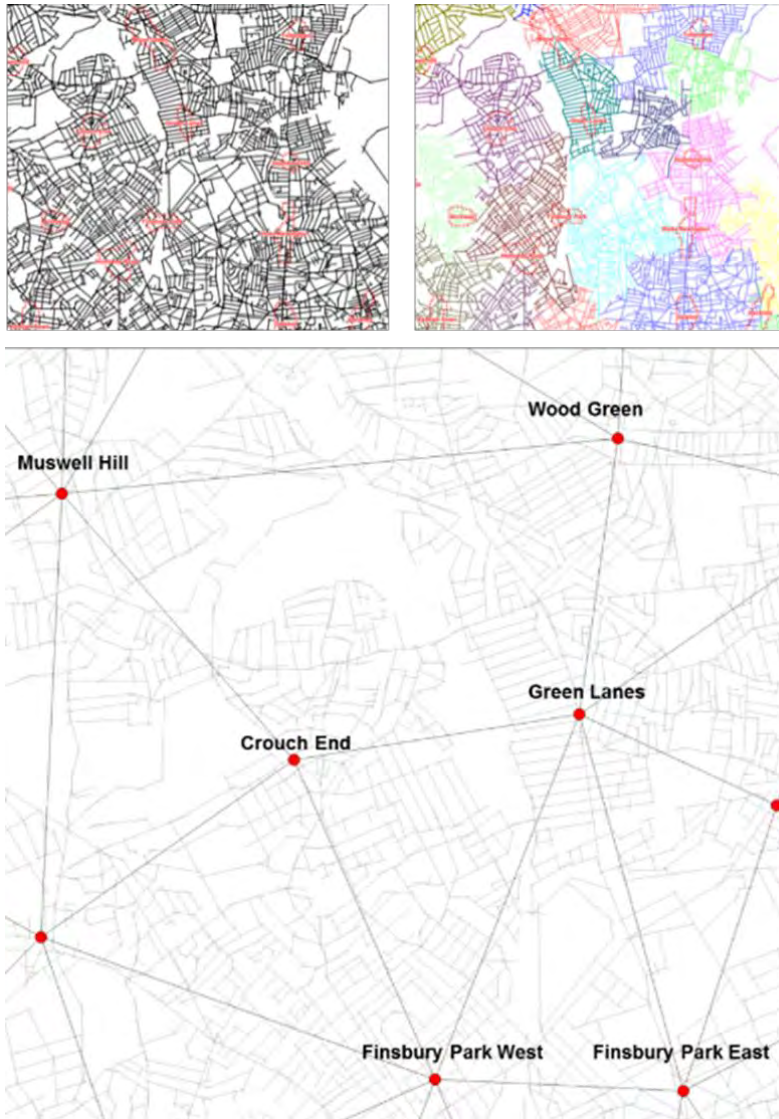


Figure 6.4 The steps in constructing the St-LA subgraph.

6.3 Empirical Strategy

In order to test the extent to which the St-LA effect is significant when correlating with house price, a number of empirical tests were undertaken. First, the homogeneity of intra-St-LA house price variations was tested with simple statistical tests. Second, the associations between inter-St-LA house price variations were analysed with a bivariate model. Third, the St-LA effect on house prices was examined with a multilevel hedonic regression model. Lastly, the effects were confirmed by comparing the St-LA with three other commonly used administrative region-based local area units by using the same multilevel hedonic regression model.

6.3.1 Intra-cluster House Price Analysis

This research conjectures that house prices are more similar within a cluster than between clusters, both visually and statistically. A one-way analysis of variance (ANOVA) was employed to test where the house price variation between the St-LAs differs from the variation within the St-LAs. The null hypothesis was that the mean house price of the sample was the same for all of the St-LAs. In the ANOVA, the F-test statistics is calculated by dividing the between group variance (B_{ms}) by the unexplained variance (W_{ms}). The null hypothesis is rejected if the p-value from the statistical test is less than 0.05.

$$F = \frac{B_{ms}}{W_{ms}}$$

where

B_{ms} is the between group variance

W_{ms} is the unexplained variance

Equation 6.4: ANOVA F-test statistics

6.3.2 Inter-cluster House Price Analysis

This research conjectures that house price homogeneity is greater between St-LAs with a greater number of connections than local areas with fewer connections. Using the St-LA subgraph network constructed in Section 3.2, this study will examine the bivariate association between inter-cluster house price variations and inter-cluster connectivity, both visually and quantitatively, for an area in North London. Figure 6.5 highlights the study area, which consists of Crouch End, Green Lanes, Finsbury Park East, Finsbury Park West, Muswell Hill and Wood Green.



Figure 6.5 North London subgraph network.

To study the relationship between inter-cluster house price variations and inter-cluster connectivity, this study first calculated the average house price per square metre in 2011 for each St-LA. House

price deviation was then calculated as the difference in absolute price per square metre between each local area and its neighbours.

$$HPDev_{ij} = |HP_i - HP_j|$$

where

HP_i is the average house price per square metre for subgraph i

HP_j is the average house price per square metre for neighbour of i

j = N(i)

Equation 6.5

The house price deviations for each St-LA and its neighbouring St-LAs were then plotted against the connectivity, the metric distance and the angular distance between the St-LAs. Statistical analysis shows the extent to which inter-cluster house price deviations were statistically associated with inter-cluster connectivity, inter-cluster metric distance and inter-cluster angular distance. More formally, these three models were calculated as follows: the house price deviation between the local area i and the adjacent local area j was regressed against the connectivity between local areas i and j, the metric distance between the geometric centroid of local areas i and j and the angular distance between the centroid of local areas i and j³⁰. Ordinary least squares (OLS) regression results, such as R-squared and p-value significance, are reported in the empirical section.

$$HPDev(i)_{ij} = B_0 + B_1 * Vij + \varepsilon$$

$$HPDev(i)_{ij} = B_0 + B_1 * Dist_{ij} + \varepsilon$$

$$HPDev(i)_{ij} = B_0 + B_1 * Ang_{ij} + \varepsilon$$

where

B₁ is the coefficient

V_{ij} is the connectivity between subgraphs

Dist_{ij} is the Euclidean distance between subgraphs

Ang_{ij} is the angular distance between subgraphs

Equation 6.6

6.3.3 Multilevel Hedonic Price Approach

Following the exploratory data analysis, this third section adopted the multilevel hedonic regression approach, which was introduced by Jones and Bullen (1994), Orford (1999) and Goldstein (1987), to estimate the effects of St-LAs on house prices in London. The multilevel hedonic regression model was chosen over a typical multiple variable OLS hedonic regression model because it examines hierarchically nested group effects. Simple OLS models ignore average variations between groups, whereas individual regressions between each local area would face sampling problems and poor generalisation. An example of a multilevel hedonic study includes the aforementioned study by Orford (1999), who provided the evidence to use multilevel models in capturing the hierarchical nature of

³⁰ One alternative to the centroid distances is to use the mean metric or angular distances between all streets of an St-LA_i to all other streets of an St-LA_j.

housing markets. Orford (1999) found that the house price variations from the grand mean were able to be decomposed into variations across enumeration districts, local communities and individual properties. Empirically, the multilevel method was also able to account for spatial autocorrelation³¹ of the error term, since properties in local areas were more similar to each other than to properties in other areas.

The following section will describe the multilevel hedonic model used in this study to model the property effect at level 1 and the local area effect at level 2³² (Stata 2012). The submarket effect at level 3 will be modelled in the next chapter. In a typical multilevel hedonic price modelling framework, the observed variable is a function of two components, namely a fixed part and a random part. The fixed part can be the mean or a collection of independent variables, and the random part is simply the deviation from the mean. To account for the hierarchical local area effects, the fixed part can be decomposed into its mean μ and broke down the random part into the local area effect μ_j , and its error e_{ijk} , as detailed below:

$$HP = \underbrace{\mu}_{\text{Fixed part}} + \underbrace{\mu_j + \epsilon_{ij}}_{\text{Random part}}$$

Where

HP is the observed house prices

μ is the mean

μ_j is the local area random effect

μ_k is the submarket random effect

e_{ijk} is the individual error term

Equation 6.7: multilevel regression model

For the empirical study, we first estimated a base grand mean model, then four nested multilevel models for the St-LAs. When local area effects were included, the dimension of the data increased. As a result, we estimated Model 3 and Model 4 with a small set of fixed predictors, namely size and accessibility, and for Model 5 a wider set of predictors. We then repeated Models 2-5 for the three region-based local areas. This includes wards in Models 6-9, LSOAs in Models 10-13, and MSOAs in Models 14-17.

³¹ "Autocorrelation is to be expected in hierarchical data, and the multilevel approach exploit this dependence" (Orford 1999 pp.7)

³² Xtmixed is the function used to estimate the multi-level model. For details please see Stata (2012).

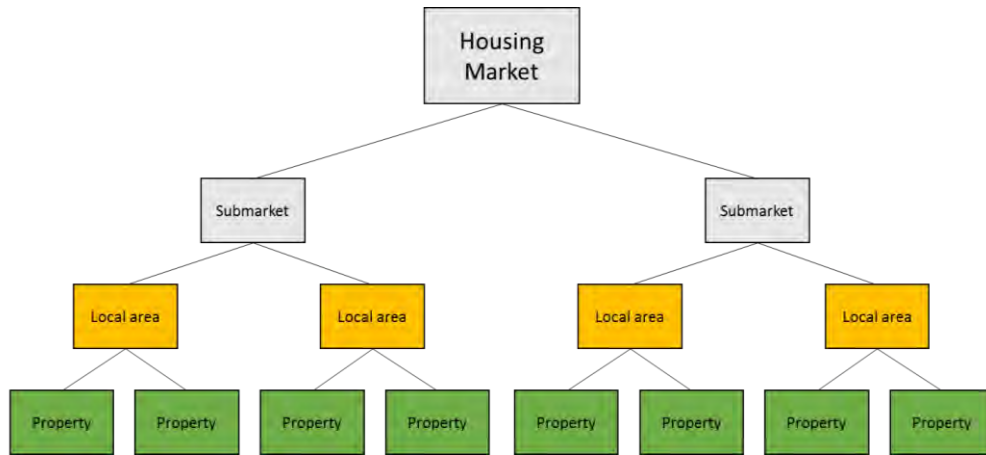


Figure 6.6 Multilevel regression framework.

Table 6.1 Candidate models.

Base Model	Model 1	Grand mean model level 1
St-LA Model	Model 2	Varying-intercept model level 2
	Model 3	Fixed predictors and varying-intercept model level 2
	Model 4	Fixed predictors and varying-intercept and varying-slope model level 2
	Model 5	Wider set of fixed predictors and varying-intercept and varying-slope model level 2
Ward	Model 6	Varying-intercept model level 2
	Model 7	Fixed predictors and varying-intercept model level 2
	Model 8	Fixed predictors and varying-intercept and varying-slope model level 2
	Model 9	Wider set of fixed predictors and varying-intercept and varying-slope model level 2
LSOA	Model 10	Varying-intercept model level 2
	Model 11	Fixed predictors and varying-intercept model level 2
	Model 12	Fixed predictors and varying-intercept and varying-slope model level 2
	Model 13	Wider set of fixed predictors and varying-intercept and varying-slope model level 2
MSOA	Model 14	Varying-intercept model level 2
	Model 15	Fixed predictors and varying-intercept model level 2
	Model 16	Fixed predictors and varying-intercept and varying-slope model level 2
	Model 17	Wider set of fixed predictors and varying-intercept and varying-slope model level 2

The starting point of the multilevel hedonic model was the base model, where no explanatory variables were specified in the regression model. This was also known as the grand mean model.

$$\text{Log}(HP_{ij}) = \mu + \varepsilon_{ij}$$

where

μ is the overall mean

ε_{ij} is the error

Equation 6.8: Model 1

Model 2 was a level two varying-intercept model that accounted for the local area effect. No explanatory variable was specified for the model.

$$\text{Log}(HP_{ij}) = \mu + u_j + \varepsilon_{ij}$$

where

μ is the overall mean

u_j denotes the local area effects

ε_{ij} is the error term

Equation 6.9: Model 2, the varying-intercept model

Model 3 was a level two varying-intercept model with fixed predictors. The predictors included space syntax integration (access) and the floor size (Floor) of the property.

$$\text{Log}(HP_{ij}) = \mu + \beta_1 * \text{Access}_i + \beta_2 * \text{Floor}_i + u_j + \varepsilon_{ij}$$

where

β_1 is the coefficient for accessibility

Access is the accessibility variable

β_2 is the coefficient for the floor size

Floor is the floor size variable

μ is the overall mean

u_j denotes the local area effects

ε_{ij} is the error term

Equation 6.10: Model 3, varying-intercept model with fixed predictors

Model 4 was a level two varying-intercept and varying-slope model with fixed predictors. The model accounted for the local area effect, adjusted for fixed effect predictors. This model included space syntax integration as both a property effect and a local area effect, which improved the statistical fit of the model.

$$\text{Log}(HP_{ij}) = \mu + \beta_{1j} * \text{Access}_i + \beta_{2j} * \text{Floor}_i + u_j + \varepsilon_{ij}$$

where

β_{1j} is the coefficient for accessibility

Access is the accessibility variable

β_{2j} is the coefficient for floor size

Floor is the floor size variable

μ is the overall mean

u_j denotes the local area effects

ε_{ij} is the error term

Equation 6.11: Model 4, varying-intercept, varying-slope model with fixed predictors

Model 5 was a level two varying-intercept and varying-slope model with a wider set of fixed predictors. This model was the same as the previous model but with the addition of the wider set of parameters,

including the dwelling type, the number of shops in the vicinity³³ and the quality of education in the vicinity³⁴.

$$\text{Log}(HP_{ij}) = \mu + \beta_1 * \text{Access}_i + \beta_2 * \text{Floor}_i + \beta_3 * \text{Dwelling1}_i + \beta_4 * \text{Dwelling2}_i + \beta_5 * \text{Shop}_i + \beta_6 * \text{School}_i + u_j + \varepsilon_{ij}$$

where

B are the coefficients for predictors

Access is the accessibility variable

Floor is the floor size variable

Dwelling1 is dummy for flats

Dwelling2 is dummy for terrace

Shop is for number of shops within 800 metres

School is the average A-level score within 800 metres

u is the overall mean

u_j denotes the local area effects

e_{ij} is the error term

Equation 6.12: Model 5, varying-intercept, varying-slope model with fixed predictors

The multilevel model was estimated using a maximum likelihood estimation (MLE)³⁵. Standard statistics for multilevel models, such as the likelihood ratio (LR), the intraclass correlation coefficient (ICC) and the Akaike Information Criterion (AIC) are reported. The LR is a test statistic that compares the goodness of fit of each candidate model with its respective null model. The test statistic is chi-square distributed and was calculated to test the significance of the local area effect on house price.³⁶ The null model is rejected in favour of the multilevel model if the p-value > 0.05. In each case, the null model was the same as the OLS model, except without the local area effect. This exclusion allowed for the isolation of the local area effect for each multilevel model. The ICC, on the other hand, is calculated for each St-LA multilevel model to measure the amount of variation captured by the local area effect in proportion to the overall house price variance.³⁷

³³ Active use is classified under the retail category in Valuation Office Agency's business rates data. Data provided by the Valuation Office Agency contains public sector information licensed under the Open Government Licence v1.0.

³⁴ A-Level scores (General Certificate of Education Advanced Level) is an academic qualification offered by educational institutions in England, Wales and Northern Ireland to students completing secondary or pre-university education

³⁵ MLE have been estimated using the Stata software which uses the Newton-Raphson gradient-based method.

³⁶ Log likelihood ratio is a common statistical test for MLE to compare goodness of fit between the null model and alternate model. The test statistic has an approximate chi-squared distribution with the degree of freedom equal to the df of alternative model – df of null model. It is calculated as follow.

$LR = -2 * [\ln(LL_{null})] + 2 * [\ln(LL_{multilevel})]$

³⁷ $ICC = \frac{(Var_{Level\ 2})^2}{[(Var_{Level\ 2})^2 + (Var_{Level\ 1})^2]}$

Var = variance

In order to compare the statistical fit across the five candidate models, the AIC was computed. The AIC³⁸ is a goodness of fit metric and a robust statistic; it will be calculated for all of the candidate models for comparison, where the lower the criterion, the better the quality of the model.

6.3.4 Model Comparison: Street-based Local Area and Region-based Administrative Local Area

This section compares the extent to which the St-LA effect differs from the region-based administrative local area effect when associated with house prices. The same multilevel hedonic approach specified in Section 3.2 was applied to three commonly used administrative units in the UK. These include electoral wards for Models 6-9, lower super output areas (LSOAs) for Models 10-13 and medium super output areas (MSOAs) for Models 14-17. Similar to the last section, the candidate models were compared through the AIC. In total, 17 candidate models were compared.

6.4 Datasets and Study Area

6.4.1 Greater London Area

As in the previous chapter, the Greater London Area in the UK was used as the case study. The extent of the study area is presented in Figure 6.7, where the black line indicates the study boundary, the red line denotes the 33 administrative borough boundaries of Greater London (ONS, 2014) and the grey line signifies the meridian line street network.



Figure 6.7 Study area boundary.

³⁸ $AIC = -2 * LL + 2 * k$

LL = loglikelihood

k = number of parameters

6.4.2 Residential Sold Price

This study used the 2011 house price dataset from the Nationwide Building Society.³⁹ House price, in this research, was defined as the exchange value between the buyer and seller. In total, 5,344 observations from 2011 were used. Figure 6.8 shows house prices in London for 2011 mapped in GIS, where red indicates a higher house price and blue indicates a lower house price. The thematic distribution in GIS was calculated by using the natural break method for eight bands.

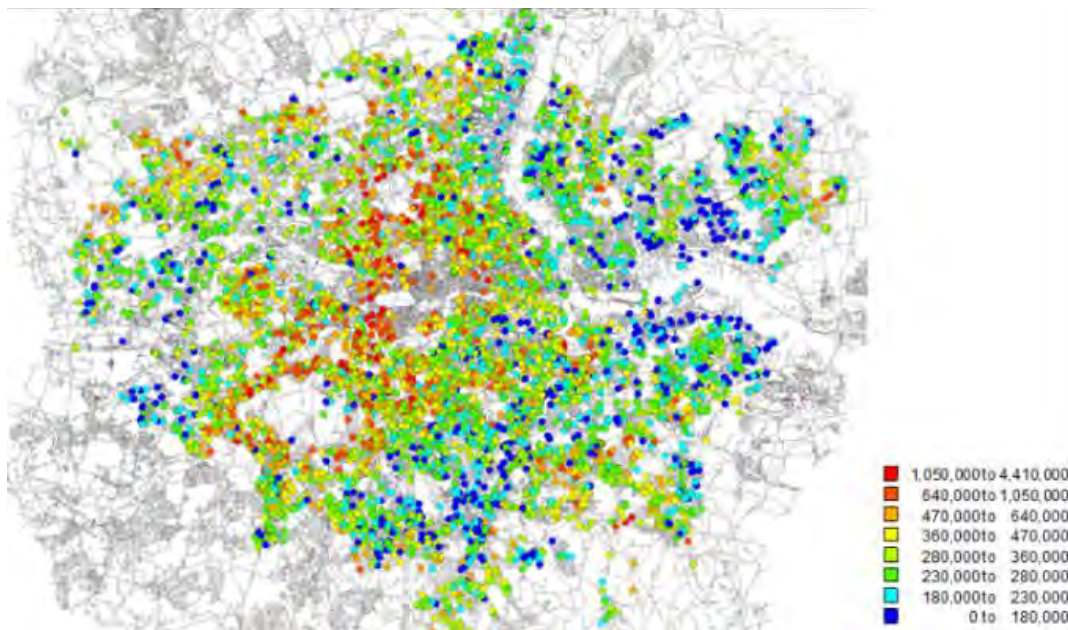


Figure 6.8 Visualisation of London house prices in 2011, with red indicating high house prices and blue indicating low house prices.

6.4.3 London Street Network

The London pedestrian street network was used to compute the accessibility measure and to construct the St-LA for the empirical study. The basis of the London street network is the Ordnance Survey (OS) Meridian street network⁴⁰ (Ordnance Survey, 2014). The spatial network dataset had a total of 113,555 street segments, as illustrated in Figure 6.9.

³⁹ The data was provided by the Nationwide through a licensing agreement with London School of Economics. The Nationwide dataset is a subset of the Land Registry dataset. The origins of all data on sold house prices in United Kingdom is owned by Land Registry/Registers of Scotland © Crown copyright 2013.

⁴⁰ Ordnance Survey Open Data Meridian 2 Dataset. © Crown Copyright {2014}



Figure 6.9 The London OS Meridian street network.

6.4.4 London Street-based Local Area

Applying the multi-level modularity algorithm, which was described in Section 2.4, on the OS Meridian line network identified a total of 207 spatial network local areas for the Greater London area. Each St-LA has an average of 549 segments, with a standard deviation of 257 segments. Table 6.2 below summarises the statistics.

Table 6.2 St-LAs detected in Greater London.

Number of Segments	
113,555	
Number of Local Areas	
207	

Segments per Local Area	
Average	549
Standard Deviation	257
Minimum	73
Maximum	1,243

Figure 6.10 below presents the St-LAs obtained from applying the multilevel modularity optimisation method to the London Meridian line map. The figure showed distinct St-LAs mapped in GIS, where the various colours corresponded to different groupings. Visually, the results displayed a clear topologic distinction for the St-LAs separated by the River Thames, such as the Isle of Dogs, and the St-LAs separated by the Lea Valley and railway tracks. This result thus presented a limitation to the method, as some areas might have been considered as continuous rather than discrete. This issue is discussed in the final section.



Figure 6.10 Visualisation of St-LA membership in the Greater London Area.

6.4.5 London Administrative Local Area Units

Table 6.3 below describes the three administrative local area units to be compared with the St-LA in the following empirical study. The smallest is the LSOA level, followed by the MSOA level and Ward level⁴¹.

Table 6.3 Local area statistics.

	N
LSOA	4,765
MSOA	983
Ward	629

Figure 6.11 illustrates the Isle of Dogs area overlaid with the LSOA (cyan), MSOA (blue) and Electoral wards (red) boundaries; the dark grey regions represent the built form. As shown above, the three boundaries entirely followed the separation created by the River Thames.



Figure 6.11 Isle of Dogs area as denoted by LSOA, MSOA and ward boundaries.

However, the divisions were more arbitrary in the central area, as boundary lines cut across streets and buildings. In contrast, the St-LA level, which was illustrated in the previous section, largely traced both the spatial separation caused by the River Thames and the morphology of the local area.

⁴¹ Electoral wards/divisions are the key local area unit for UK administrative geography. There are a total of 9,456 wards in the UK with an average population of 5,500 people per ward. (ONS 2015)

6.4.6 Descriptive Statistics

Table 6.4 describes the set of variables included in the study. Similar to the last chapter, this set includes structural features such as property size; dwelling type (e.g. flat, house or terrace); location accessibility, such as street network closeness centrality; and neighbourhood amenities, like the number of retail units within 800 metres (Law et al., 2013; Des Rosier et al., 1996) and the secondary school average scores within 800 metres (Black, 1999; Gibbons and Machin, 2003; 2008). The mean house price is approximately £350,000, with an average floor size of 99 square metres, an average 2.6 bedrooms and a mean property age of 85 years old.

Table 6.4 Descriptive statistics for house prices and attributes.

Variables	Descriptions	(1) Mean	(2) Standard Deviation
Price	Transaction Price	356,481	213,846
Bedrooms	Number of Bedrooms	2.604	1.006
Floors	Floor Size	99.03	40.73
Age	Age of Property	85.05	36.35
CC	Closeness Centrality	8,721	2,719
BC	Betweenness Centrality	2.643e+07	1.087e+08
Shops	Number of Shops within 800 metres	354.3	407.7
Parks	Distance to Parks and Gardens	10,355	5,272
Schools	Average A-level Score within 800 metres	366.2	390.6
new_build_dum1	More than Five Years Old	0.986	0.117
new_build_dum2	Newly Built	0.0139	0.117
tenure_dum1	Freehold	0.592	0.492
tenure_dum2	Leasehold	0.408	0.492
type_dum1	Terrace	0.314	0.464
type_dum2	Flat	0.405	0.491
type_dum3	House	0.281	0.449

(For details of the individual dataset, please see Chapter 4.)

6.5 Empirical Results

The following section illustrates the empirical results for this chapter. First, it examines whether house price variations between St-LAs differ when compared to house price variations within St-LAs, as specified in Section 3.1. Next, the relationship between inter-cluster house price variation and inter-cluster connectivity is assessed, as specified in Section 3.2. Then, using the multilevel hedonic approach as specified in Section 3.3, the significance of the house prices is tested. Last, a comparison is made between the associations of the St-LAs and the traditional administrative local areas with house prices, as specified in Section 3.4.

6.5.1 Intra-cluster House Price Analysis Results

In this section, the intra-cluster house price variations within each St-LA, in comparison with the between-cluster variations, are examined both visually and quantitatively. House price variations were first visually examined for an area in North London, and then, with a one-way ANOVA, scrutinised for the entire city. The one-way ANOVA applied an F-test to determine whether the house price variations between the St-LAs differed from the within-St-LA variations.

Table 6.5 illustrates the ANOVA results, which tested whether the 2011 house price variations of the St-LAs differed from the within variations. The p-value was statistically significant, at a 0.01 level. These initial results showed, quantitatively, house prices in London were significantly more similar within each St-LA than between.

Table 6.5 ANOVA Statistics. The results suggest house prices are more similar within local areas.

	Sum of Squares	Df	MS	F	Prob > F
Between Groups	4.20E+15	165	2.54E+13	121.35	0
Within Groups	1.46E+16	69487	2.10E+11		
Total	1.88E+16	69652	2.69E+11		

Figure 6.12a illustrates the community detection membership for an area in North London, and Figure 6.12b presents the house prices per square metre for the same area. The four circled areas were from the western part of North London, including Highgate, Crouch Hill, Green Lanes and Seven Sisters. The results showed greater house price homogeneity within the St-LAs than between. The most obvious was the difference between Crouch Hill and Green Lanes, which only had two connections between them due to the spatial separation of the railway tracks. The difference between Highgate and Crouch Hill was much less than between Green Lanes and Crouch Hill, which suggested that the house price heterogeneity differed. This heterogeneity between local areas is explored in the next section.

Community Detection



House Price

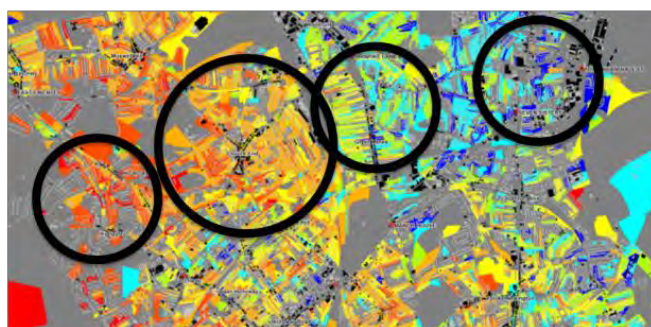


Figure 6.12 London spatial house price clustering effect.

a. Community detection memberships in North London. Colours denote different memberships.

b. The 2011 house prices in North London. Red denotes higher house prices whilst blue denotes lower house prices.

6.5.2 Inter-cluster House Price Analysis Results

In this section, the same area in North London was used to examine, both visually and quantitatively, the relationship between inter-cluster house price variations and inter-cluster connectivity. Table 6 describes these six areas, each local area's adjacent local areas, the house price deviation between them, its connectivity, and the metric and angular distances between them.

Table 6.6 North London St-LA inter-cluster house prices.

Local Area	Adjacent Local Area	Price Difference in 2011	Connections	Metric	Angular
Crouch End	Finsbury Park	0.09	16	2,590.59	2.72
	Highgate	0.08	12	2,090.81	2.51
	Muswell Hill	0.15	6	2,303.41	2.03
	Green Lanes	0.26	2	2,516.26	3.05

Local Area	Adjacent Local Area	Price Difference in 2011	Connections	Metric	Angular
Green Lanes	Wood Green	0.20	11	2,457.66	3.19
	Tottenham	0.29	8	3,296.66	2.72
	Finsbury Park East	0.43	3	1,963.69	0.61
	Haringey-Seven Sisters	0.25	3	1,493.24	1.54
	Finsbury Park West	0.23	2	2,721.80	1.51
	Crouch End	0.36	2	2,266.14	3.35

Local Area	Adjacent Local Area	Price Difference in 2011	Connections	Metric	Angular
Finsbury Park East	Dalston	0.05	21	2,567.24	2.30
	Stamford Hill - Stoke Newington	0.20	10	2,138.05	2.82
	Finsbury Park West	0.14	10	2,253.32	2.01
	Angel Islington	0.17	9	2,746.19	3.36
	Green Lanes	0.30	3	2,227.61	1.33
	Haringey-Seven Sisters	0.47	2	2,141.62	1.63

Local Area	Adjacent Local Area	Price Difference in 2011	Connections	Metric	Angular
Finsbury Park West	Crouch End	0.10	16	2,623.45	2.54
	Lower Holloway	0.08	11	2,993.97	1.82
	Finsbury Park East	0.16	10	1,880.25	1.57
	Angel, Islington	0.35	10	2,545.35	2.19
	Highgate	0.19	8	2,757.16	1.80
	Kentish Town	0.20	6	1,991.22	2.43
	Green Lanes	0.19	2	2,782.06	1.78

Local Area	Adjacent Local Area	Price Difference in 2011	Connections	Metric	Angular
Wood Green	Green Lanes	0.24	11	1,978.55	1.65
	Palmers Green	0.07	11	2,713.01	3.56
	Tottenham	0.12	9	2,993.51	1.83
	Bounds Green	0.11	4	2,588.80	2.65
	Muswell Hill	0.44	1	2,948.85	3.32

Local Area	Adjacent Local Area	Price Difference in 2011	Connections	Metric	Angular
Muswell Hill	Bounds Green	0.23	8	3,320.64	2.14
	Crouch End	0.17	6	1,856.70	2.11
	Hampstead Garden Suburb - East Finchley	0.16	4	3,384.18	5.70
	Finchley	0.12	4	4,220.65	3.70
	North Finchley	0.15	2	4,367.34	3.95
	Highgate	0.26	1	2,733.08	4.49
	Wood Green	0.30	1	3,569.53	3.54
	unknown	0.07	1	2,514.46	4.10

Figure 6.13 illustrates the St-LA subgraph for the study area, where the nodes are coloured according to the 2011 average house prices per square metre, and the thickness of the edge denotes the number of connections between the St-LAs. For example, the St-LA designated as Crouch End had an average house price per square metre of approximately £5,000. The adjacent St-LA designated as Green Lanes had an average house price per square metre of approximately £3,700. The house price

deviation is the absolute difference between the two, which, in this case, was £2,300. The poor connectivity between these two local areas was reflected in the large house price differences; this will be seen in the descriptive analysis in the last section.



Figure 6.13 The visualised London St-LA subgraph map.

To explore these results, Figure 6.14 presents three scatterplots. On the Y-axis are the St-LA house price deviations between the St-LA and the St-LA's neighbours; the St-LA's corresponding connectivity and metric and angular distances between the local areas are on the X-axis. The first two scatterplots showed a poor association between the house price deviations and the adjacent local areas' metric and angular distances. The third scatterplot showed a significant negative relationship between the house price deviations and the subgraph connectivity.



Figure 6.14 Scatterplot between inter-cluster house price deviation and inter-cluster distances.

To confirm this association, three regression models were constructed; the results are reported in Table 6.7. Model 1 regressed the log of the house price differences with the log of the metric distance between the local areas and attained an R-squared of 3.6%. The p-value showed that the metric distance impedance between the local areas is insignificant. Model 2 regressed the log of the price differences with the log of the angular distance between the local areas and realised a slightly improved R-squared of 5.3%. The p-value showed that the angular distance impedance between the

local areas was insignificant. Model 3 regressed the log of the price differences with the log of the connectivity between the local areas and achieved an R-squared of 50.4%. The p-value showed that the connectivity between the local areas was significant.

The results showed that the variations in the inter-cluster house price deviations could be partially explained by the inter-cluster connectivity between the local areas. The low association with the distance was expected, as the St-LAs were mostly equidistant. These results revealed that the transaction price buyers were willing to pay was not only related to the local area the property sat in, but also to which local areas it was connected and how strong this connection was. This result showed that buyers were not simply purchasing a part of Crouch End but were, perhaps, also purchasing a location closer to Highgate than to Green Lanes. This evidence showed the importance of associating spatial configuration with house prices at the subgraph level rather than only at the street network level. The next section uses the multilevel regression framework to study the local area effects on house prices whilst controlling for different predictor variables.

Table 6.7 Inter-cluster regression analysis results.

Variables	(1) Model 1	(2) Model 2	(3) Model 3
ln_metric	-0.0574 (0.0579)		
ln_ang		-0.0680 (0.0562)	
ln_connections			-0.0656*** (0.0128)
Constant	1.239** (0.451)	0.890*** (0.0814)	0.933*** (0.0283)
Observations	28	28	28
R-squared	0.036	0.053	0.504

Standard errors in parentheses.

*** p < 0.01, ** p < 0.05, * p < 0.1

6.5.3 Street-based Local Area Multilevel Regression Results

The first part of this analysis studies the extent to which St-LA effects are evident when associated with house price variations, as specified in Section 3.1. Table 6.8 below illustrates the regression results for the five candidate models.

Table 6.8 Multilevel hedonic regression results.

Street-based Local Area	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed					
Integration			0.046*** (0.008)	0.058*** (0.013)	0.034*** (0.012)
Floor Size			0.342*** (0.003)	0.340*** (0.003)	0.277*** (0.003)
Age					0.030*** (0.003)
Park					0.127*** (0.013)
Shops					0.034*** (0.005)
Terrace					0.093*** (0.004)
Flat					0.030*** (0.003)
School					0.010*** (0.003)
_cons	12.660*** (0.006)	12.620*** (0.022)	12.620*** (0.020)	12.610*** (0.020)	12.630*** (0.018)
Random					
Var(Residual)	0.220 (0.004)	0.151 (0.003)	0.043 (0.001)	0.041 (0.001)	0.034 (0.001)
St-LA					
var (cons)		0.085 (0.010)	0.078 (0.009)	0.070 (0.009)	0.058 (0.007)
var (integration)				0.016 (0.00297)	0.014 (0.00245)
N	5,334	5,334	5,334	5,334	5,334
Loglikelihood	-3,536	-2,781	477	542	1022
LR test					
Prob > chi2	0***	0***	0***	0***	0***

*** p<0.01, ** p<0.05, *p<0.1

Model 1 was the null model, where the grand mean of the log house price was 12.660; the residual illustrates the total variance away from the mean was 0.220. Model 2 was a level two varying-intercept model, where the between-St-LA (level 2) variance in house prices was 0.085 and the between-property (level 1) variance was 0.151. Model 3 was a level two varying-intercept model with fixed predictors, where the between-St-LA (level 2) variance in house prices was 0.078 and the between-property variance was 0.043. This reduction in the property variance was to be expected, due to the predictor inclusion. Model 4 was a level two varying-intercept and varying-slope model with fixed predictors, where the between-St-LA (level 2) variance in house prices was 0.070, the between-

St-LA integration (level 2) variance was 0.016, and the between-property (level 1) variance in house prices was 0.041. Model 5 was a level 2 varying-intercept and varying-slope model with wider sets of fixed predictors, where the between-St-LA (level 2) variance in house prices was 0.058, the between-St-LA integration (level 2) variance was 0.014, and the within-property (level 1) variance in house prices was 0.034. This reduction in the property variance was, again, expected due to the wider set of fixed predictors.

The local area effect remains relatively stable, with a small reduction due to the inclusion of the fixed effect predictors in Model 3 and Model 5. This result demonstrated the relative stability of the local area effect on house prices. The overall loglikelihood ratio test (Prob > chi-squared = 0***) showed significance for all of the candidate models, which presented robust evidence that the St-LA effect was significant. In terms of estimates, this research found that 5-8% of house price variations can be explained by the St-LA effect.

Table 6.9 below summarises the goodness of fit, as measured by the AIC, between the five candidate St-LA multilevel models. The reduction in the AIC showed a clear improvement of the statistical significance, which allowed for the progressive inclusion of the local area effect and the fixed predictor effect. In addition, the local area effect in Model 2 and the fixed predictor in Model 3 and Model 5 had significant improvements in statistical significance. These improvements indicated that the housing market was hierarchically nested for at least two levels, namely the property level and the local area level.

Table 6.9 Comparison of the AICs.

Street-based Local Area	Model 1	Model 2	Model 3	Model 4	Model 5
Akaike Information Criterion (AIC)	7,075.495	5,567.136	-944.593	-1,072.618	-2,040.18

Figure 6.15 below summarises the intra-class correlation coefficient for each model. The ICC measured the amount of variation captured by the local area effect and the property effect, in proportion to the overall house price variance. Blue denotes property variance, and orange denotes local area variance. The local area variance was consistently above 30% and remained relatively

stable after the inclusion of the fixed predictors.

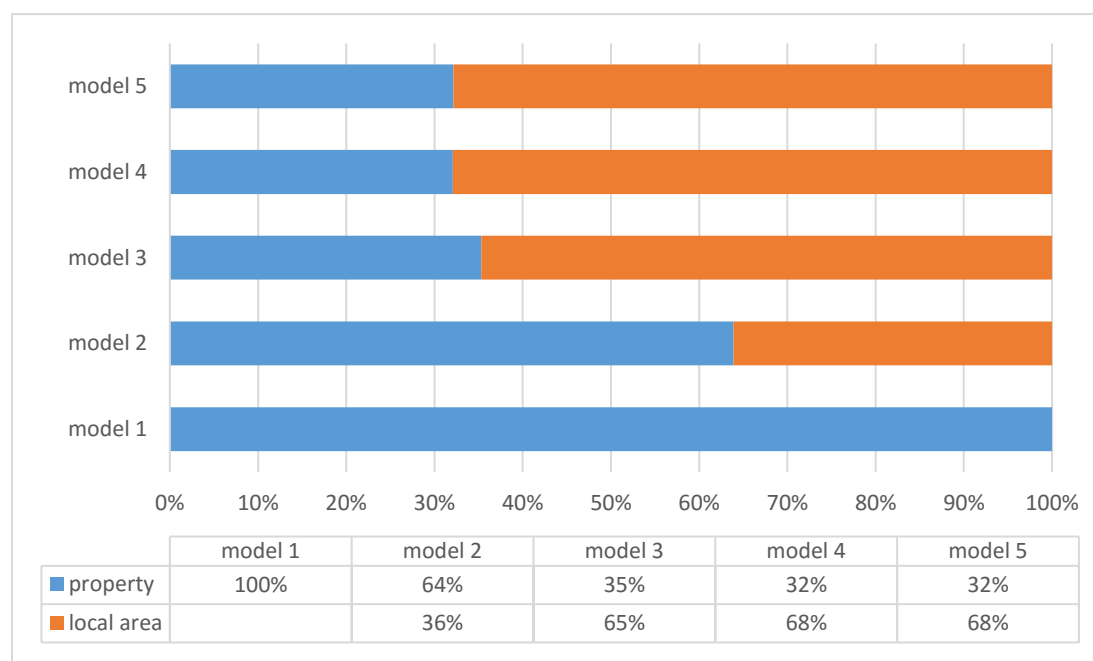


Figure 6.15 Intra-class correlation coefficient comparisons.

For empirical reasons, spatial autocorrelation effects were checked. For Model 1 and Model 5, Global Moran's I was calculated with a minimum radius of 2,400 metres. This radius was used to ensure there was a significant sample for each data-point to calculate the statistic. The global spatial autocorrelation reduced from 0.27 in Model 1 to 0.004 in Model 5. The p-value showed a weak significance, with prob > 0.01. This result confirmed the previous research on the use of multilevel hedonic models in reducing spatial autocorrelation (Orford, 1999). For details, please see appendix C.

6.5.4 Comparison of Street-based Local Areas and Administrative Local Areas

This section compares the five candidate models using different local area units. The LR tests for all five candidate models were significant. Figure 6.16 below shows a goodness of fit comparison between the five local area units using the AIC. The St-LA is denoted by the colour blue, followed by ward in orange, the MSOA in grey and the LSOA in yellow.

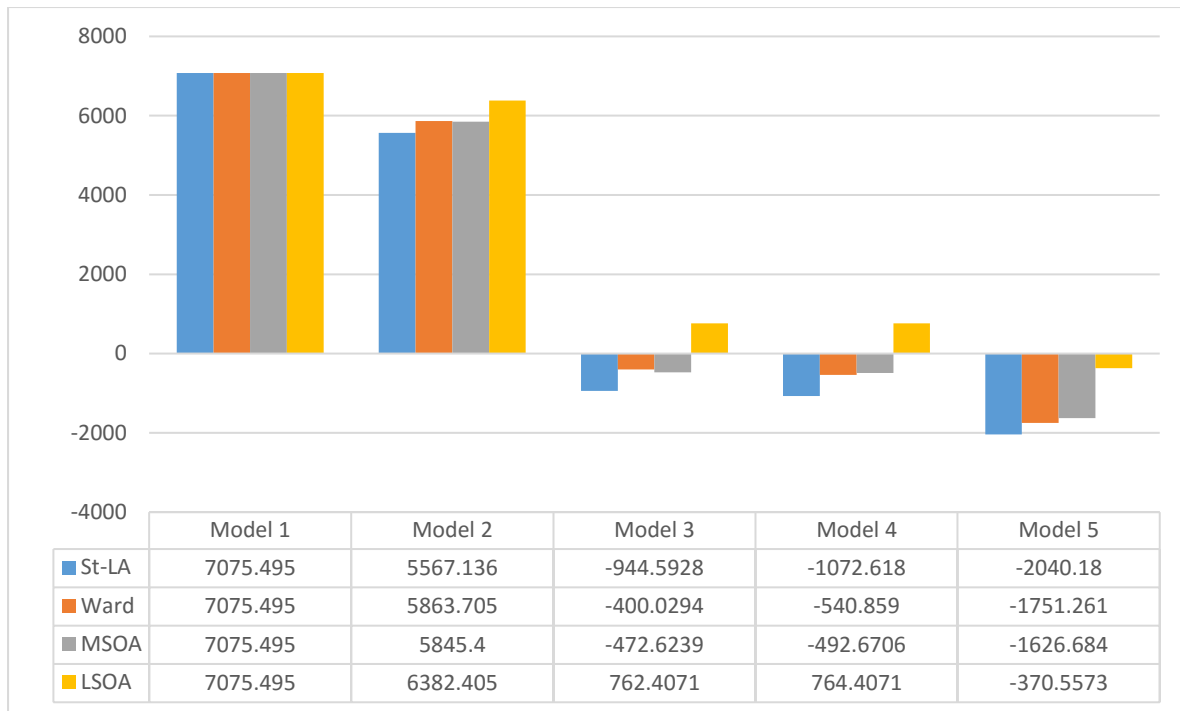


Figure 6.16 Goodness of fit comparison (AIC) across different local area units.

The downward trend of the AIC showed the progressive, joint effect of the property characteristics and the local area effects on house prices across the local area units. This result confirmed Orford's research (1999) on the hierarchical nature of the housing market, where the London housing market was nested in at least two levels. This outcome also showed clear differences in the results across different administrative units. The results also demonstrated that St-LAs are consistently preferred to all the other administrative units, including electoral wards, MSOAs and LSOAs. Together, this evidence confirmed the effect of St-LAs on house prices.

6.6 Discussion

This research applied community detection techniques to define the St-LAs in London. Multilevel hedonic model results show that the local areas have significant effects on house prices and that the St-LAs are preferred to the administrative units. Interpretations of these key findings are illustrated below.

The main contribution of the research is the novel application of the community detection techniques on the street network dual graph in order to define the St-LAs in London. The results show that local areas have significant effects on house prices and that the St-LAs are able to capture socio-economic similarities more accurately than region-based local area units. The plausible reasons are threefold. First, people perceive the local area on a street network; therefore, the street network is able to more precisely capture subtle differences in an urban environment and more accurately the perceptual definition of a local area than ad-hoc administrative regions. Figure 6.17 explains this concept, which illustrates two distinct local areas (one in orange and one in green) connected by a bridge (grey).

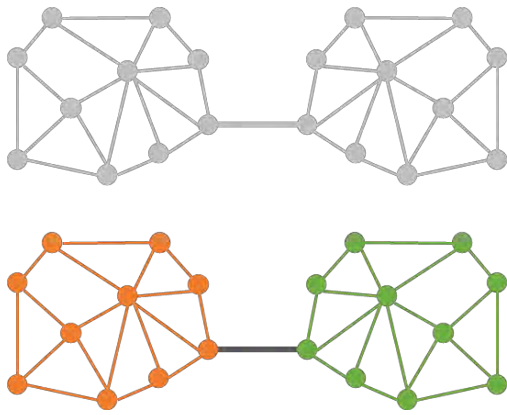


Figure 6.17 Two abstract subgraph connected via a bridge.

If we randomly pick any orange node in the network, the chance of arriving at another orange node is much greater than the possibility of reaching a green node. Using this analogy, the probability of walking within the same subgraph or identifying the highly connected subgraph as a local area are much greater than in another subgraph. On aggregate, the topology of the street network can more accurately capture the perceptual definition of the local area. This result can provide a linkage between spatial network clusters and the collective perception of neighbourhoods. To verify this, future empirical research is needed, where individual perception maps are compared to St-LA units (Coulton et al. 2001).

Secondly, the topology of the street network reinforces socio-economic similarities within the local area, and over time reinforces the perceptions of the local area. Figure 6.18 illustrates this, where we run a simple simulation of an agent walking on the same spatial network.⁴²

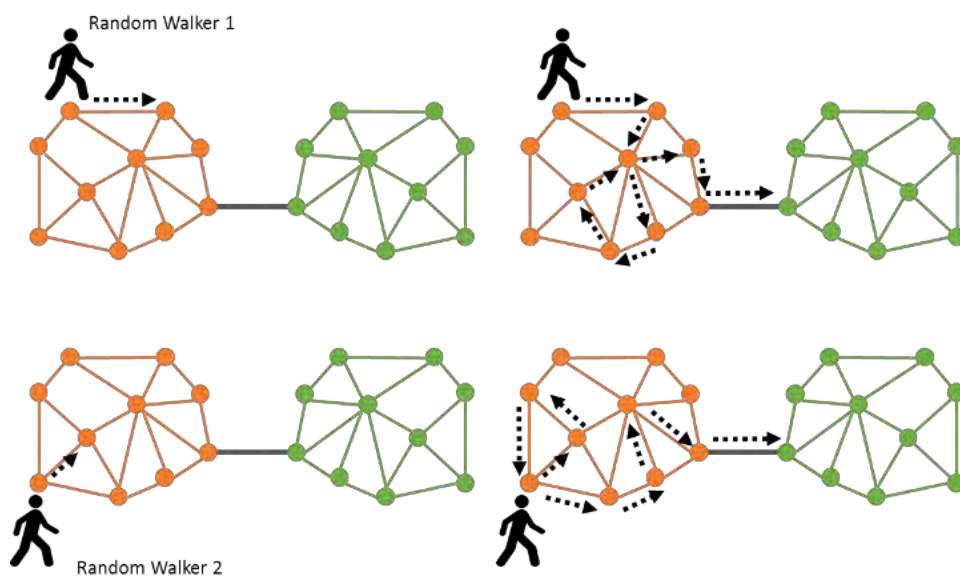


Figure 6.18 Simulation of a random walker on the abstract graph.

⁴² This notional simulation takes inspiration from the Walktrap algorithm (Pons and Latapy, 2006).

The simulation begins with an agent who starts from a random orange node then arbitrarily walks to a connected node. The number of steps required to reach the green subgraph is then recorded. The first simulation in Figure 18 shows that the agent took nine steps to reach a green node. The second simulation in the figure shows that the agent took eight steps to reach a green node. A plausible future scenario is that, over time, differences between areas will become more pronounced, as like-minded people cluster together and bump into each other. This result, therefore, reinforces socio-economic similarities within an St-LA and the boundaries between St-LAs. Plausible processes allowing this to happen include crowd behaviour and bounded rationality where information is constrained within the local area (Benerjee, 1992; Simon, 1957). To verify this, a key future question to ask is, 'To what extent do social constructs, perceptual clusters and topologic clusters overlap with St-LAs across space and time?'

Thirdly, as people identify these local areas, the local area becomes part of the housing bundle, leading to an effect on the house prices. For example, when an individual purchases a property in Kensington, they are also buying a Kensington local area premium as part of the housing bundle. Therefore, a buyer will value a house more similarly to another within the same local area than a house in another local area. From the geographical science perspective, this can also be interpreted by using Tobler's first law, where properties that are closer to each other are likely to be more socio-economically similar than properties that are further apart (Tobler, 1970). Over time, local areas will become more socio-economically homogeneous, further reinforcing the effect on the house prices.

6.6.1 Benefits and Limitations

There are a number of benefits in defining St-LAs. First, St-LAs can more accurately capture subtle differences in urban environments than ad-hoc administrative regions. Second, as the street network is clearly the most permanent of all morphological elements, St-LAs can be considered as a slow dynamic. This slowness allows the data to be consistently compared across time, though at the same time being dynamic enough to reflect the changes in the street network and the morphology. To demonstrate the benefits, the simulation described in Figure 6.18 was repeated 500 times. Figure 6.19 illustrates the average number of random steps (500-runs) an agent at an orange subgraph would need to take to reach the green subgraph for four different configurations.

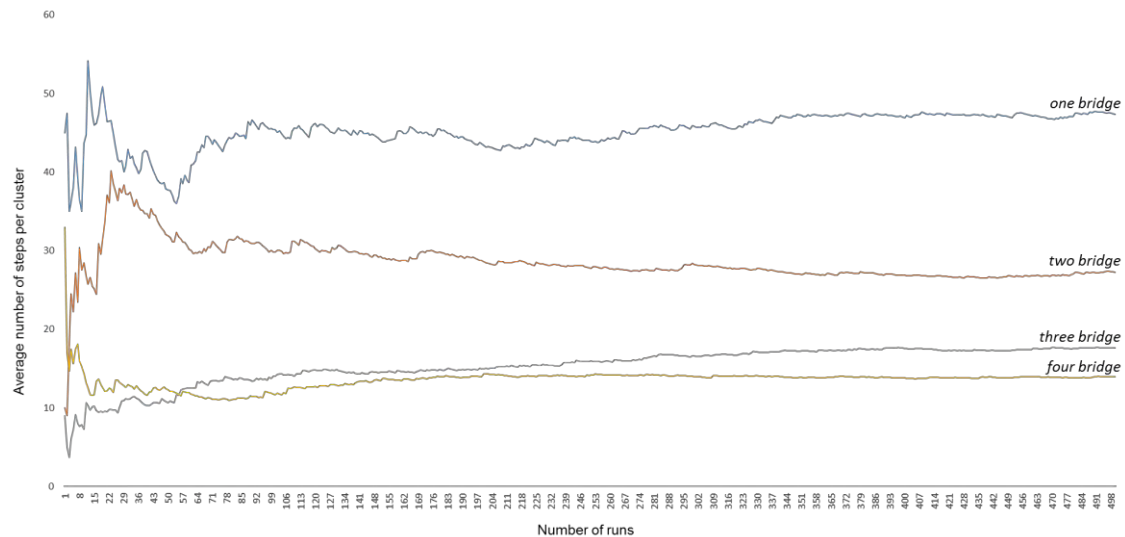


Figure 6.19 Average number of steps required to jump between clusters.

For the agent to cross one bridge, it takes an average of 40 random steps; to cross two bridges, it takes an average of 30 steps; to cross three bridges, it takes an average of 15 steps; and, to cross four bridges, it takes an average of 10 steps. One can see that the more bridges there are between the two subgraphs, the lower the number of average random steps any agent will need to take in order to reach the adjacent subgraph. The simulation shows that if we add one more bridge across the two St-LAs, the probability of arriving at a green node increases substantially. The result shows how subtle difference of the spatial street network can lead to great differences in terms of co-presence. This in effect can change the perception of local area boundary and socio-economic spillover.

The definition of St-LA is not without its concerns. First, this research suggested that, on aggregate, the St-LAs were able to capture subtle differences in an urban environment more accurately than region-based methods. However, at an individual level, more research is required to understand and confirm how this happened and to determine the processes that influenced the construction of the individual's neighbourhood boundaries (Tolman 1948). Second, consideration of the street network provided a one-dimensional approach to defining a local area. When a grid is highly uniform and connected, street network connectivity might not be an important factor in identifying neighbourhoods. For example, in many American CBDs, the grid is too uniform to be separated; instead, these areas might be more defined by other dimensions such as morphological, sociological, economical and historical characteristics. For example, a local area can be identified from its density such as Midtown Manhattan. Basing neighbourhoods on a single spatial variable reduces its feasibility for use in spatial planning. Future research is recommended to focus on joining multiple factors in order to create a more comprehensive definition of local areas or neighbourhoods for planning.

Third, the use of the multilevel modularity optimisation method defined sharp local area boundaries, which contradicted previous research, where neighbourhoods were described as fuzzy and

overlapping (Alexander 1964). To overcome this limitation, future studies can apply fuzzy-logic memberships in community detection. Lastly, research is needed to examine how St-LAs can improve existing housing research topics, such as the definition of housing submarkets. This topic is discussed in the next chapter, which focuses on housing submarket formation.

6.6.2 Conclusion

Despite the limitations of this approach, the definition of St-LAs is important, as it links the geometry of the street network to the way we perceive the local area. This research provides evidence that the configuration of the spatial network should be considered when specifying the local area definition. For real estate economists, this research highlights the local area effects on house prices, which are important in house price prediction models. For urban planners, this research reveals considerable evidence that neighbourhoods are not only defined by socio-economic or historic dimensions, but also through their spatial network topology or configuration. This conclusion is important, as administrative census tracts have been used in many aspects of spatial planning. Street-based methods can, therefore, provide an alternative to ad-hoc administrative local geographies for neighbourhood planning and policies.

Chapter 7

Street-based Housing Submarket

7.1 Introduction

Over the past few decades, many housing studies have been conducted on the topic of housing submarkets (Adair et al., 1996; Leishman, 2009; Bourassa 2002; MacLennan and Tu, 1996; Dale-Johnson, 1982; Bourassa et al., 1999; Schnare and Struyk, 1976; Watkins, 2001). For some submarkets, having better access to different social opportunities matters more, while for other submarkets, better access to a good school might be more important. These submarkets are related to demographics, housing policies and service provision. Thus, a better understanding of the housing submarket can create more informed housing policy; ignoring these processes can result in inefficiency and poor resource allocation. A key objective of this chapter is to extend from the previous chapter in applying St-LAs to defining Street-based Housing Submarkets (St-HSs). This research argues that St-HSs have a significant effect on house prices and that they also provide a stronger statistical fit than a traditional housing submarket formed by an administrative local area. This research applies the multilevel hedonic price model to identify housing submarket effects on house prices. Figure 7.1 illustrates the housing submarket focus for this chapter.

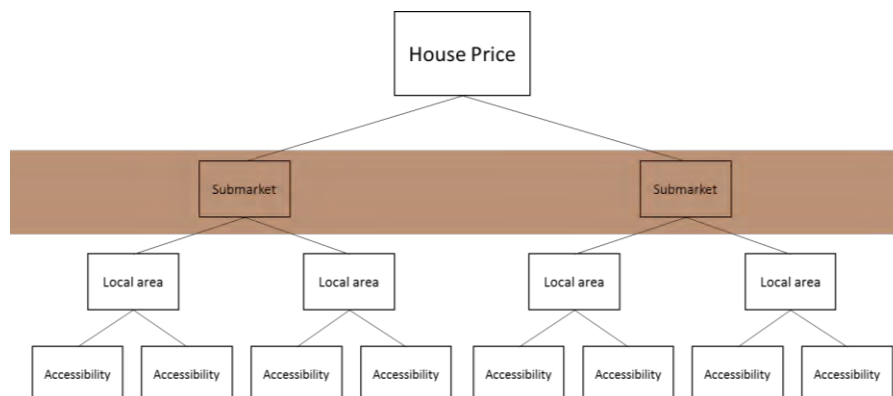


Figure 7.1 This chapter focuses on the housing submarket level.

The remainder of the chapter is organised as follows: Section 1 provides the research background, a short overview of housing submarkets and defines the research question. Section 2 describes the framework, St-HSs and procedures used to construct housing submarkets. Section 3 provides details on the empirical method used in the study. Section 4 introduces the thesis case study. Section 5 reports the regression results, and Section 6 provides a general discussion of the findings.

7.1.1 Housing Submarkets

Rosen (1974) described the underlying economic model for a composite good, such as housing, which is made up of utility-bearing parts. In equilibrium, the market will settle on a set of clearing prices. However, the market's clearing prices are not expected to equalise across property markets, as properties are inherently unique. This uniqueness brings about inefficiency in the market, where supply and demand do not instantly equalise. This inefficiency or market disequilibria can be caused by the

time lag from inefficient housing supply, information asymmetry from high search costs or from differences in demand across socio-economic groups. The differences in implicit prices across local markets have given rise to the housing submarket concept, which is one of the most discussed topics in housing studies (Maclennan and Tu, 1996; Leishman, 2009).

Grisby et al. (1987) defined housing submarkets as units that are reasonable substitutes for one another but relatively poor substitutes for units in other submarkets. This homogeneity can be determined a priori, such as the use of real estate agents (Bourassa, 2002), or empirically driven, such as through the combination of spatial and structural factors (Allen et al., 1995; Strasheim, 1975). Common ways to segment markets in housing studies include structural factors, such as dwelling type, e.g. whether the residence is a flat or detached home (Schnare and Struyk, 1976; Bajic, 1985; Adair et al., 1996; Allen et al., 1995). Another way to segment markets in housing studies is through spatial factors. For example, if a home is in a particular school catchment area or a neighbourhood (Strasheim, 1975; Galster, 1987). Oftentimes, both spatial and structural dimensions are clustered together to form housing submarkets (Watkins 2001; Adair et al. 1996). This bundling has naturally led to the use of statistical techniques in grouping similar properties across multiple factors (Bourassa, 1999), which is possible due to improvements in computation that allow the statistical clustering of higher dimension data. Many studies combine structural features, such as dwelling type, and location factors, such as school catchments, into an empirically driven spatial-structural housing submarket. The motivation for this is that housing submarkets are not simply the construct of a building type or a neighbourhood, but rather a combination of all of these attributes. This is logical, as a buyer who wants to buy a detached house might also want to be near a park with suitable access to a primary school.

7.1.2 Strand Three Research Question Definition

Despite consensus on the existence of housing submarkets and which statistical test to use, there are general disagreements concerning the methods utilised to identify these submarkets and which variables to include (Watkins, 2001; Schnare and Struyk 1976; Bourassa et al., 1999). One topic that is rarely discussed in the literature is the geography used to construct housing submarkets. The general procedure is that traditional administrative local areas, such as census tract or wards, are aggregated into housing submarkets through a statistical clustering procedure. However, these 'arbitrary' or 'ad hoc' local area units can create problems in the identification of housing submarkets. For example, Goodman and Thibodeau (2002) compared the results between a school zone aggregated submarket, the census block aggregated submarkets and the zip code aggregated submarkets. The results show significant differences across local area units, which suggests the inconsistency problem in the housing submarket definition can partly be attributed to local area units. This irregularity in geography is related to the modifiable areal unit problem, which describes a source of statistical bias when data points are aggregated (Openshaw, 1983). What is clearly missing in the literature is the need to better understand how local areas affect the housing submarket definition in hedonic price models. This assertion leads to the next section, which proposes the concept of the St-HSs. The analytical chapter focuses on the following research question:

Research strand three question: To what extent are St-HSs comparable to traditional census tract-based housing submarkets when correlated with house prices?

7.2 Street-Based Housing Submarket Framework

The previous chapter conjectured that St-LAs have significant effects on house prices in London. This chapter expands upon this notion by proposing that St-HSs also have significant effects when correlated with house prices and that they are preferred to traditional housing submarkets. Explaining this preference is beyond the scope of this study. Instead, this research intends to provide evidence that St-LAs can be used to improve both the housing submarket definition and the goodness of fit in a hedonic price model. A plausible explanation is that, in comparison to administrative units, St-LAs can more accurately capture perceived local area boundaries, as individuals experience the urban environment at the street level. As a result, St-LAs might encapsulate spatial housing submarket definitions more accurately than traditional administrative local areas. Extending from the previous chapter, this chapter suggests that St-LAs improve housing submarket effects on house prices. To study the research question, this chapter compares housing submarkets formed by St-LAs and housing submarkets shaped by traditional administrative boundaries in a hedonic price model. Figure 7.2 shows this comparison, where the diagram on the left illustrates a housing submarket that combines spatial, structural and network factors, and the diagram on the right illustrates a housing submarket that uses spatial and structural factors does not consider St-LA factors.

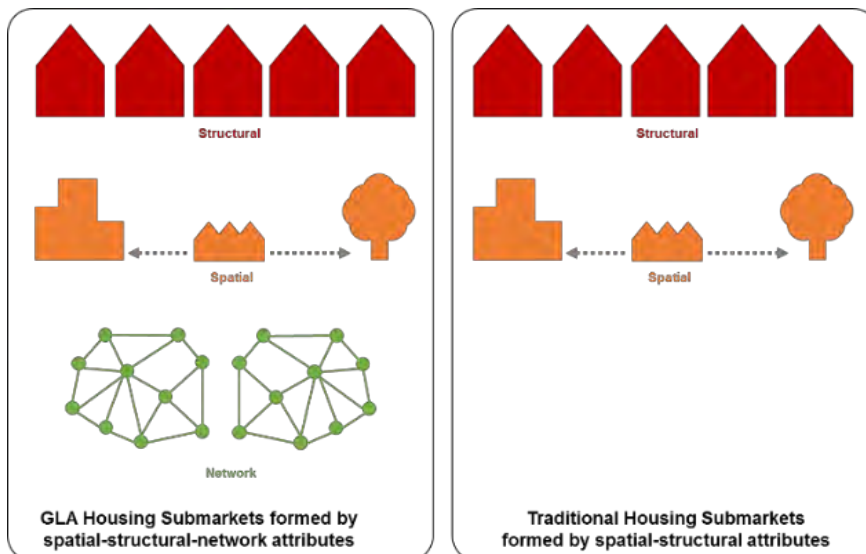


Figure 7.2 Street-based housing submarkets that combine street network attributes, spatial attributes and structural attributes (left) compared to traditional housing submarkets formed by spatial-structural attributes (right).

7.2.1 Submarket Construction Procedures

The submarket construction procedure is based on a three-step process. First, select a geography or

locality. Second, calculate the averages of the structural and location characteristics within each local area. Third, employ a statistical classification method for each local area average to define k-housing submarkets. This research will construct and compare the postcode unit submarket, the LSOA submarket, the MSOA submarket, the ward submarket and the street-based submarket. Figure 7.3 illustrates these procedures.

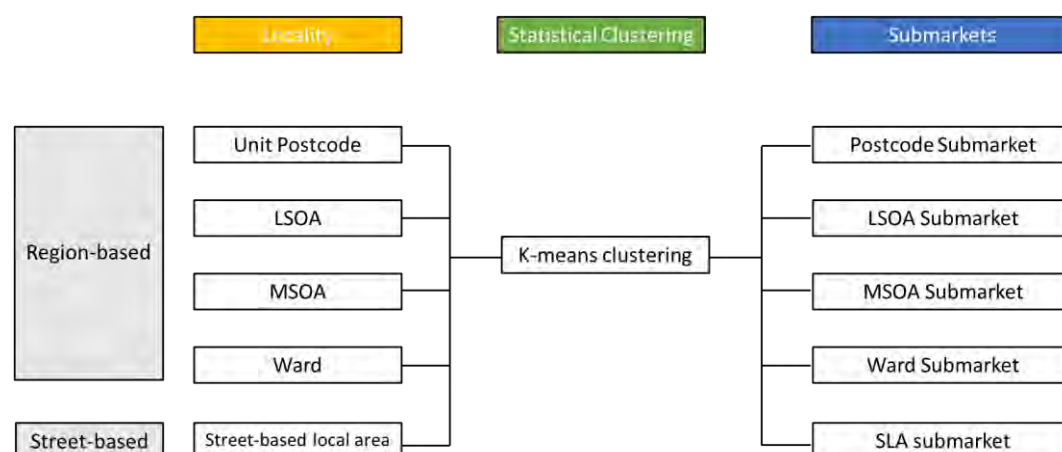


Figure 7.3 The submarket construction process.

7.2.2 Unsupervised K-means Clustering Algorithm

A number of clustering methods have previously been employed to define housing submarkets that encompasses multi-dimensions. These techniques often include k-means clustering (GLA 2004; Leishman 2009; Day et al. 2002), hierarchical clustering (Goodman and Thibodeau 1998), PCA (Bourassa et al. 1999) or the more complex machine learning methods such as artificial neural network classification methods (Kaoko 2002). Dale-Johnson (1982), for example, used factor analysis on structural attributes to demarcate housing submarkets in Santa Clara. Goodman and Thibodeau (1998) used hierarchical clustering to define five housing submarkets in Dallas. Bourassa et al. (1999) used PCA to reduce dimensionality and statistical clustering on housing attributes to delineate five housing submarket clusters in Sydney and Melbourne. Day et al. (2002) used principal component analysis, k-means clustering and hierarchical clustering to define four housing submarkets in Glasgow. The Greater London Authority (2004) used k-means clustering techniques on socio-economic and housing characteristics to define five to six housing submarkets in London.

This study will adopt one of the most standard techniques, the unsupervised k-means clustering technique (GLA 2004; MacQueen, 1967; Bourassa, 1999), which aims to partition n observations into k clusters that minimise the differences between seven property attributes, namely the dwelling type, bedrooms, the floor size, the dwelling age, space syntax integration, park amenities and school amenities. K-means clustering (Equation 7.1) was discussed briefly in the methodology chapter.

$$argmin_s \sum_{l=1}^k \sum_{x \in S_l} \|x - \mu_l\|^2$$

x = values for factor

u = mean for factor

Equation 7.1

The reason for the use of this method is that it has been more widely adopted than newer methods, such as the machine learning classification method or the fuzzy logic clustering method. This method is also more suitable than hierarchical linear models, which can render different numbers of clusters. This research uses standard variables, such as the dwelling type, the size, the accessibility and the amenities to construct the housing submarket. The employment of k-means clustering on a standard set of variables allows us to focus on comparing the local areas (traditional vs street-based) rather than across clustering methods and variables. Future research can examine how local area definition might influence different types of housing submarket (eg. income, socio-demographics).

7.3 Empirical Method

This research adopted the hedonic price approach for the analytical study (Orford 1999; Schnare and Struyk 1976). Two separate analyses were used. First, this research tested the significance of the St-HSs through the widely adopted hedonic submarket test (Schnare and Struyk 1976). This test ensured that implicit prices were similar within and different between housing submarkets in order to satisfy the housing submarket condition (Chapter 3). Second, this research examined the effects of St-HSs on house prices through a multilevel hedonic price regression model (Orford, 1999).

- Housing submarket tests
- Multilevel hedonic price regression model

7.3.1 Housing Submarket Test

The housing submarket test (Schnare and Struyk 1976) is a standard procedure used to test the significance of housing submarkets. The main premise of this approach is that implicit prices are significantly different across housing submarkets. The housing market is first partitioned into

$k = \{2, \dots, 10\}$ submarkets, taking the averages of the following property attributes within each local area or geography via the k-means clustering algorithm: the floor area, the number of bedrooms, the binary factor on flats and houses, the property's age, space syntax integration or closeness centrality, average A-level score within 800 metres and the distance to parks and gardens. Second, a simple OLS regression model (7.2) was estimated for each housing submarket, where house price was regressed against a vector of independent variables using an OLS estimator.

$$\text{Log}(P_i) = \beta_1 X_i + e_i$$

Where

P_i is the price of a property for i = postcode

X_i represents a vector of independent variables

B are the coefficients for the independent variables

e_i is the error term

Equation 7.2

The standard OLS specification included the typical dwelling structural variables, amenity-specific variables and location-specific variables. In general, a property with more space, bedrooms, amenities and accessibility was expected to render a higher house price. Third, a standard statistical test, known as the F-test, was used to examine whether there were significant differences between any pairs of the submarket regression equations under the null hypothesis, which assumed that the two models were equivalent. This was tested for every combination of the St-HSs, where $k = \{2, \dots, 10\}$.⁴³ Lastly, a weighted standard error test was applied to find the optimal number of clusters needed to reduce the greatest proportion of weighted standard errors (WSE)⁴⁴. The threshold for the WSE was set to 10%. The specification for the chosen St-HS model is reported and described below.

7.3.2 Multilevel Hedonic Price Regression Model

This research also adopted the multilevel hedonic price regression model, introduced by Jones and Bullen (1994) and Orford (1999), to estimate the St-HS effect on house prices in London. The rationale for using this model over a typical OLS model is that it examines nested group effects. The simple OLS model ignores the average variations between groups, whereas individual regression models between each submarket face sampling problems and poor generalisation. Orford (1999; 2001) showed evidence for using multilevel hedonic price models to capture the hierarchical nature of housing submarkets. He found that house price variations from the grand mean can be deconstructed into variations across enumeration districts, local communities and individual properties. Orford (1999) suggests multilevel methods are also able to account for spatial autocorrelation effects⁴⁵ of the error term, because properties in local areas are more similar to each other than to properties in other areas.

This study adopted the multilevel framework to estimate both the nested local area and submarket effects on house prices (Stata 2012⁴⁶). The following section describes the multilevel hedonic price model used for this study to model the property effect at level one, the local area effect at level two and the submarket effect at level three. In a typical multilevel hedonic price model framework, the observed variable is a function of two components, namely a fixed part and a random part (Equation 7.3). The fixed portion can be the mean or a collection of independent variables, and the random part is simply

⁴³ The test statistic is given by where; SSR_1 , SSR_2 and $SSRC$ are the sum of squared residuals for the individual models and the combined model and N_1 , N_2 and K_1 , K_2 are the number of observations and number of parameters in the individual models respectively. The test statistic, F^* , has an F distribution with $\text{Min}(K_1, K_2)$, $(N_1 + N_2) - (K_1 + K_2)$ degrees of freedom.

$$F = \frac{(SSR_c - (SSR_1 + SSR_2))}{(SSR_1 + SSR_2)} * \frac{((N_1 + N_2) - (K_1 + K_2))}{\text{Min}(K_1, K_2)}$$

SSR = sum of squared residuals

⁴⁴ The formula for the standard error test is as follows where N_j is the number of transactions in the j th submarket, k_j is the number of explanatory variables in the j th submarket equation.

$$SE_c = \frac{(N_1 - K_1 - 1)}{\sum(N_j - K_k - 1)} * SE_1 + \dots + \frac{(N_j - K_k - 1)}{\sum(N_j - K_k - 1)} * SE_j$$

SE = Standard Error

⁴⁵ "Autocorrelation is to be expected in hierarchical data, and the multilevel approach exploit this dependence" (Orford 1999 pp.7)

⁴⁶ `Xtmixed` is the function used to estimate the multi-level model. For details please see Stata (2012).

the deviation from the mean. To account for the hierarchical local area and submarket effects, we deconstructed the fixed segment into its mean μ and broke down the random part into the local area effect μ_j , submarket effect μ_k and its error ϵ_{ijk} , as detailed below:

$$HP = \underbrace{\mu}_{\text{Fixed part}} + \underbrace{\mu_j + \mu_k + \epsilon_{ijk}}_{\text{Random part}}$$

Where

HP is the observed house prices

μ is the mean

μ_j is the local area random effect

μ_k is the submarket random effect

ϵ_{ijk} is the individual error term

Equation 7.3: Multilevel regression model.

For the empirical study, we first estimated a base grand mean model and then five nested multilevel models. We estimated Models 2, 3, 4 and 5 with a parsimonious set of fixed predictors, namely floor area and accessibility, and for Model 6 a wider set of predictors, including the number of shops, the quality of the primary and secondary schools, the distance to the parks and the dwelling type, as discussed in the previous chapter. Figure 7.4 illustrates the framework for the multilevel hedonic model, and Table 7.1 depicts the first six candidate models to be estimated.

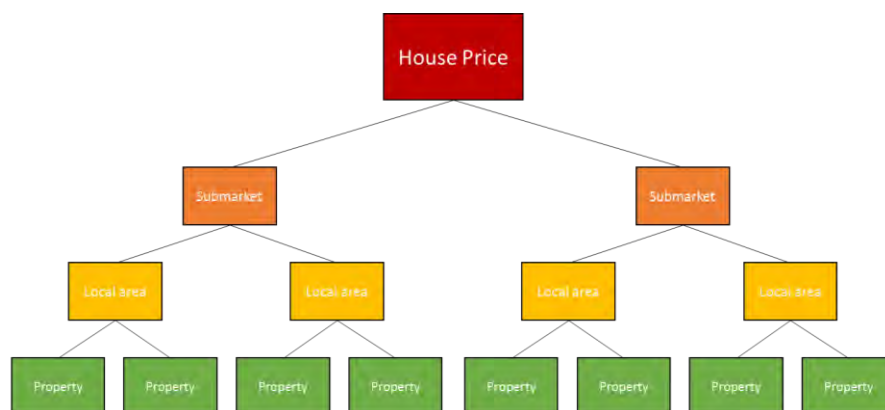


Figure 7.4 Multilevel hedonic price model framework

Table 7.1 The candidate models.

Base Model	Property	Model 1a	Grand mean model level one
Street-based Model	Local Area	Model 2a	Varying intercept model level two
		Model 3a	Varying intercept model with fixed predictors level two
		Model 4a	Varying intercept model and slope model with fixed predictors level two
	Housing Submarket	Model 5a	Varying intercept model and slope model with fixed predictors level three
		Model 6a	Varying intercept model and slope model with wider set of fixed predictors level three

Model 1 is the base model, also known as the grand mean model, in which there is no explanatory variable specified in the regression model (Equation 7.4).

$$\text{Log}(HP_{ijk}) = \mu + \varepsilon_{ijk}$$

Where

HP is the house price

μ is the overall mean

ε is the error term

Equation 7.4: Model 1 is the grand mean model.

Model 2 is the level two varying intercept model that accounts for local area effects. No explanatory variable is specified in the model.

$$\text{Log}(HP_{ijk}) = \mu + \mu_j + \varepsilon_{ijk}$$

Where

HP is the house price

μ is the overall mean

μ_j is the local area effect

ε_{ijk} is the error term

Equation 7.5: Model 2 is the level two varying intercept model.

Model 3 is the level two varying intercept model with predictor variables. The predictors are space syntax integration **A** and the floor size **F** of the property.

$$\text{Log}(HP_{ijk}) = \beta_1 * A_i + \beta_2 * F_i + \mu + \mu_j + \varepsilon_{ijk}$$

Where

HP is house price

μ is the overall mean

β_1 is the accessibility coefficient

A is the accessibility variable

β_2 is the floor size coefficient

F is the floor size variable

μ_j is the local area effect

ε_{ijk} is the error term

Equation 7.6: Model 3 is the level two varying intercept model with fixed predictors.

Model 4 is the level two varying intercept and slope model with fixed predictor variables. This model accounts for the local area effect, adjusted for the predictors. This model includes space syntax integration as both a property effect and a local area effect.

$$\text{Log}(HP_{ijk}) = \beta_{1j} * A_i + \beta_2 * F_i + \mu + \mu_j + \varepsilon_{ijk}$$

Where

HP is house price

μ is the overall mean

β_{1j} is the fixed and random coefficient for accessibility

A is the accessibility variable

β_2 is the coefficient for floor size

Floor is the floor size variable

μ_j is the local area effect

ε_{ij} is the error term

Equation 7.7: Model 4 is a level two varying intercept, varying coefficient model with fixed predictors.

Model 5 is the level three varying intercept and slope model with fixed predictors. This model accounts for the local area effect and the housing submarket effect, adjusted for the predictors.

$$\text{Log}(HP_{ijk}) = \beta_{1j} * A_i + \beta_{2j} * F_i + \mu + \mu_j + \mu_k + \varepsilon_{ijk}$$

Where

HP is the house price

B1 is the accessibility coefficient

A is the accessibility variable

B2 is the floor size coefficient

F is the floor size variable

u is the overall mean

u_j is the local area effect

u_k is the submarket effect

e_{ijk} is the error term

Equation 7.8: Model 5 is the level three varying intercept, varying coefficient model with fixed predictors.

Model 6 is the level three varying intercept and slope model with a wider set of predictors. This model accounts for the local area effect and the housing submarket effect, adjusted for a wider set of predictors. Model 6 is the same as Model 5 but with the additional explanatory variables, which include dwelling types *D1* and *D2*, the shop amenity variable *S* and the school amenity variable *E*.

$$\text{Log}(HP_{ijk}) = \beta_{1j} * A_i + \beta_{2j} * F_i + \beta_{3j} * D1_i + \beta_{4j} * D2_i + \beta_{5j} * S_i + \mu + \mu_j + \mu_k + \varepsilon_{ijk}$$

Where

HP is the house price

B is the coefficient for predictors

A is the accessibility variable

F is the floor size variable

D1 is the dummy for flats

D2 is the dummy for terrace

A is for number of shops within 800 metres

S is the average A-level score within 800 metres

u is the overall mean

u_j is the local area effect

u_i is the submarket effect

e_{ijk} is the error term

Equation 7.9: Model 6 is the level three varying intercept and varying coefficient model with a wider set of fixed predictors.

The multilevel models were estimated using a maximum likelihood estimator (MLE)⁴⁷. Standard statistics for multilevel models, such as the loglikelihood ratio test (LR), the Akaike information criterion (AIC) and the intra-class correlation coefficient (ICC) were reported. The LR test, a test statistic, compared the goodness of fit for each candidate model with its respective null model. The test statistic

⁴⁷ MLE have been estimated using the Newton-Raphson gradient-based method.

was Chi-square distributed and was calculated to test the significance of the local area effect on house prices.⁴⁸ The null model was rejected in favour of the multilevel model if the p-value < 0.05. The null model in each case was the same as the OLS model without the local area effect. The LR test allowed for the isolation of the local area effect for each multilevel model. The ICC, on the other hand, was calculated for each multilevel model to measure the amount of variation the local area effect and the housing submarket effect captured in proportion to the overall house price variance.⁴⁹ The *AIC*⁵⁰ is a statistical fit metric which was used to compare different models; this statistic was more robust than *r*², as it adjusted for the number of variables in the model. The *AIC* was calculated for all of the candidate models and compared.

7.3.3 Comparing Street-based and Traditional Housing Submarkets

This section compares the extent to which St-HSs and traditional housing submarkets formed by region-based administrative local areas differ. The six candidate models in Section 3.2 were applied to the St-LA (a), the statistical ward (b), the MSOA (c), the LSOA (d) and the postcode unit (e); this rendered a total of 30 candidate models, which are illustrated in Table 7.2. Like the last chapter, the candidate models were compared using the AIC goodness of fit.

⁴⁸ Log likelihood ratio is a common statistical test for MLE to compare goodness of fit between the null model and alternate model. The test statistic has an approximate chi-squared distribution with the degree of freedom equal to the df of alternative model – df of null model. It is calculated as follows.

$$LR = -2 * [\ln(LL_{null})] + 2 * [\ln(LL_{multilevel})]$$

LL = loglikelihood

$$^{49} ICC = \frac{(Var_{Level 2})^2}{[(Var_{Level 2})^2 + (Var_{Level 1})^2]}$$

Var = variance

$$^{50} AIC = -2 * LL + 2 * k$$

LL = loglikelihood

k = number of parameters

Table 7.2 Local area housing submarket models specifications.

- (a) Street-based local area
- (b) Statistical ward
- (c) Medium Super Output Area
- (d) Lower Super Output Area
- (e) Postcode unit

	Model	Description	St-LA	Ward	MSOA	LSOA	Postcode
Property	1	Grand mean model level one	1a	1b	1c	1d	1e
Local Area	2	Varying intercept model level two	2a	2b	2c	2d	2e
	3	Varying intercept model with fixed predictors level two	3a	3b	3c	3d	3e
	4	Varying intercept model and slope model with fixed predictors level two	4a	4b	4c	4d	4e
Housing Submarket	5	Varying intercept model and slope model with fixed predictors level three	5a	5b	5c	5d	5e
	6	Varying intercept model and slope model with a wider set of fixed predictors level three	6a	6b	6c	6d	6e

7.4 Dataset and Study Area

7.4.1 Greater London Area

As in the previous two chapters, the Greater London Area was used as the case study. The extent of the study area is presented in Figure 7.5, where the black line indicates the study boundary, the red line indicates the 33 administrative borough boundaries of Greater London (ONS, 2014). This chapter also used the house price dataset from Nationwide Building Society and the Land Registry;⁵¹ a total of 5,344 observations from 2011 were used.



Figure 7.5 The Greater London study area.

⁵¹ The data was provided by the Nationwide through a licensing agreement with London School of Economics. The Nationwide dataset is a subset of the Land Registry dataset. The origins of all data on sold house prices in United Kingdom is owned by Land Registry/Registers of Scotland © Crown copyright 2013.

7.4.2 London Local Area Units

Five local area units were used for the analytical study, including the St-LA, the statistical ward, the MSOA, the LSOA, and the postcode unit, as presented in Table 7.3.

Table 7.3 Local area statistics.

Local Area	N
St-LA	206
Ward	629
MSOA	983
LSOA	4,765
Postcode	197,066

Figure 7.6 illustrates these five local areas for the London Thamesmead region. All of the local area units showed division caused by the River Thames separation. Critically, the region-based administrative local area bore little relation to the local area street network or the urban morphology. While this result was evident across all of the region-based units, it was most evident for the postcode unit, the MSOA and the LSOA and less so for the electoral wards. The St-LA level, in contrast, followed both the large separation caused by the River Thames and the spatial network morphology of the Thamesmead development.



Figure 7.6 The five local area units in Thamesmead.

7.4.3 Descriptive Statistics

Below is the set of variables included in the hedonic price model, which are the same as the variables in the previous chapter. Table 7.4 describes the basic statistics for the 2011 London house price dataset. The mean house price was approximately £350,000, with an average floor size of 99 square metres, 2.6 bedrooms on average and a mean property age of 85 years old.

Table 7.4 Descriptive statistics for house prices and attributes.

Variables	Description	(1) Mean	(2) Standard Deviation
Price	Transaction Price	356,481	213,846
Bedrooms	Number of Bedrooms	2.604	1.006
Floors	Floor Size	99.03	40.73
Age	Age of Property	85.05	36.35
CC	Closeness Centrality	8,721	2,719
BC	Betweenness Centrality	2.643e+07	1.087e+08
Shops	Number of Shops within 800 m	354.3	407.7
Parks	Distance to Parks and Gardens	10,355	5,272
Schools	Average A-level score within 800 m	366.2	390.6
new_build_dum1	More than Five Years Old	0.986	0.117
new_build_dum2	New-build	0.0139	0.117
tenure_dum1	Freehold	0.592	0.492
tenure_dum2	Leasehold	0.408	0.492
type_dum1	Terrace	0.314	0.464
type_dum2	Flat	0.405	0.491
type_dum3	House	0.281	0.449

7.5 Empirical Results

The following section illustrates the empirical results for this chapter. The study first tested the existence of St-HSs (Section 3.1). Then, the significance of St-HSs on house prices was tested using the multilevel hedonic price regression model (Section 3.2). This was followed by a comparison between the St-HS and the traditional housing submarket (Section 3.3).

7.5.1 The Housing Submarket Tests

The standard housing submarket test (Schnare and Struyk 1976) was used to identify the existence and specification of the St-HS. It consisted of the F-test, which tested the extent to which the submarket pairs differed, and a WSE test, which examined the extent to which the housing submarket model improved upon the non-housing submarket model. Table 7.5 summarises the F-test for every St-HS pair. We tested nine models, where the number of submarkets equalled $K = \{2, \dots, 10\}$. The results showed that the St-HS model was significant at $p\text{-value} < 0.01$ up to $K \leq 8$, where the null hypothesis was rejected for all pairs. Above $K = 8$, there was at least one submarket pair where the null hypothesis was not rejected. This result confirmed the existence of an St-HS model where $K \leq 8$.

Table 7.5 Chow test summary (F-tests).

F-tests Summary	Significant Pairs	Insignificant Pairs
K2	1	0
K3	3	0
K4	6	0
K5	10	0
K6	15	0
K7	21	0
K8	28	0
K9	34	2
K10	48	1

Figure 7.7 summarises the WSE test for the St-HS model, where the number of submarkets was equal to $K = \{2, \dots, 8\}$. The result showed $K = 7$ and $K = 8$ reached the 10% threshold for WSE reduction. A 12% WSE reduction was achieved for $K = 7$ and $K = 8$ achieved an 11% WSE reduction. This result showed that St-HS was significant and that the most optimal partition was $K = 7$. This housing submarket specification was used for the remainder of the study.

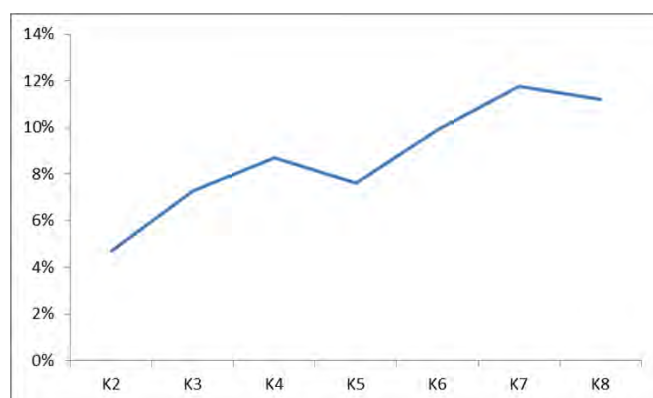


Figure 7.7 Weighted mean square error test summary.

7.5.2 London Street-Based Housing Submarkets (St-HS)

Figure 7.8 illustrates the seven St-HSs in London. The results are coloured and interpreted. Table 7.6 illustrates the characteristics of these seven housing submarkets. The first submarket was interpreted as 'the outer extension', with a mean house price of £330,000, an average of 2.6 bedrooms and mostly consisted of terraces. This submarket included Walthamstow, Richmond, Twickenham and South Norwood. The second submarket was interpreted as 'East London', which had a mean house price of £350,000, an average of 1.95 bedrooms and high levels of closeness centrality. This submarket included Whitechapel, Southwark, Canada Water and Canary Wharf. The third submarket was interpreted as 'the inner extension', with a mean house price of £420,000, an average of 2.4 bedrooms

and mostly comprised flats. This submarket covered many urban neighbourhoods, including Clapham, Dalston, Hampstead, Highgate, Brixton, Camberwell and New Cross. The fourth submarket was interpreted as 'West London', with a mean house price of £520,000, an average of 1.97 bedrooms and the highest levels of closeness centrality. This region included the central neighbourhoods of Chelsea, Kensington, Notting Hill and Pimlico. The fifth submarket was interpreted as 'the suburbs', with a mean house price of £320,000, an average of 2.9 bedrooms, low levels of closeness centrality and comprised mostly houses. This submarket included the areas of Finchley and Barnet. The sixth submarket was interpreted as 'the working class suburb', with a mean price of £240,000, an average of 2.7 bedrooms and lower levels of closeness centrality. This area included Edmonton, Southall, Sutton, Dagenham and Romford. The seventh submarket was interpreted as 'the edge of London', with a mean house price of £340,000, an average of three bedrooms and the lowest levels of closeness centrality. This submarket included the areas of Hornchurch, Orpington, West Wickham and Stanmore. The results showed clear differences in spatial-structural-network attributes across the different London housing submarkets. Submarkets in Central London generally had smaller homes with greater accessibility, while submarkets in outer London largely had larger homes with lower accessibility. This result implied a key trade-off between accessibility and space.

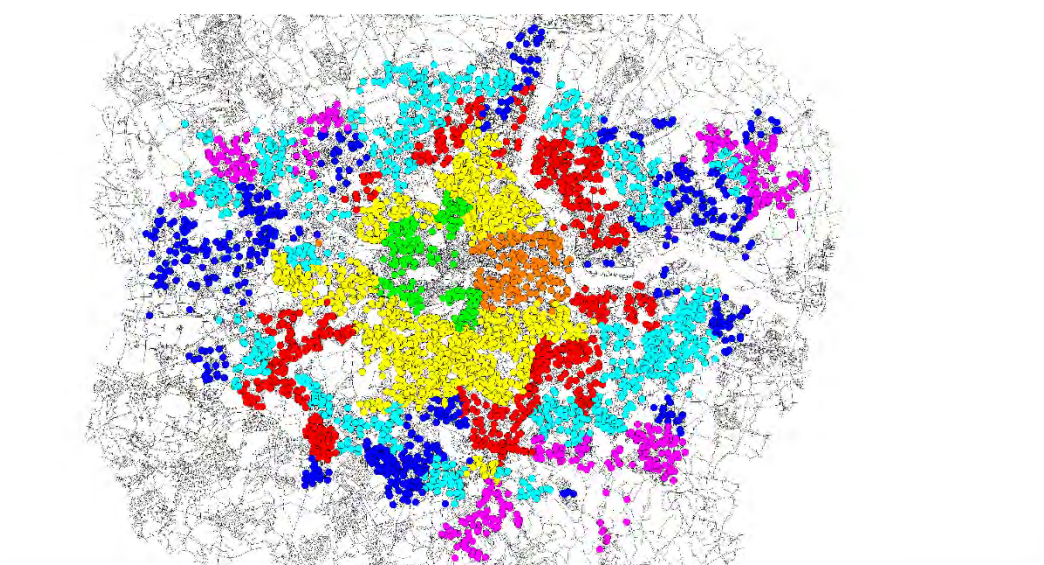


Figure 7.8 Greater London St-HSs.

Table 7.6 Greater London St-HS summary.

Classification	Interpretation	Average Price	Average Bedrooms	Dominant Dwelling Type	Closeness Centrality
	Outer Extension	330,000	2.6	Terraces	Medium High
	East London	350,000	1.95	Flats	High
	Inner Extension	420,000	2.4	Flats	Medium High
	West London	520,000	1.97	Flats	High
	Suburb	320,000	2.9	Houses	Low
	Working Class Suburb	240,000	2.7	Houses	Low
	Edge of London	340,000	3.0	Houses	Low

The results showed a clear East-West distinction in Central London, where the West had a higher-income population and the East has a lower-income population. The results also depicted the concentric pattern of London housing submarkets, with Inner London being more clustered and Outer London more fragmented. One interpretation is that traditional neighbourhoods in the centre of the city are more connected to surrounding neighbourhoods, while newer developments are more fragmented and isolated. The results also showed that large green space can be either an integrator, such as Hyde Park, or a segregator, such as the Lea Valley. One limitation of this classification is the lack of North-South distinction in the submarket analysis. This suggests that housing attributes did not differ between North and South London. One plausible explanation is that the submarket analysis did not consider public transport connectivity, which differs significantly between these areas. These results reflect the potentiality of South London.

7.5.3 London St-LA Submarkets Individual Regression Results

Table 7.7 summarises the individual regression results for each St-HS. All the coloured cells are significant at the p -value < 0.05 level and all the white cells are insignificant at the p -value < 0.05 level. The cell colour range goes from green for negative estimates to red for positive estimates.

Table 7.7 Individual housing submarket regression summary.

		Outer Extension	East London	Inner Extension	West London	Suburban London	Working Class Suburbs	Edge London
Integration								
New Build								
Tenure								
Flats								
Terrace								
Choice								
Parks								
School Quality								
Age								
Bedrooms								
Floor Size								
Shops								
Insignificant								
Significant	-							
	+							

The results showed that not all of the variables were significant for all the submarkets, reflecting the

differences in supply and demand for these attributes. Space syntax integration and the size of the property were the two variables that were significant throughout. This result is understandable for all housing submarkets and is supported by the traditional bid-rent model in modelling the trade-off between space and accessibility. As the distribution for some of the housing attributes are heterogeneous across space, differences in attribute value can reflect both a lack of demand or unequal supply of the attribute. For the two central submarkets, the tenure attribute was not significant. This result reflects the potential lack of supply of these attributes rather than the lack of demand as the tenure of the building is related to the growth of the city. More simply, there is a limited supply of freehold house for sale in Central London.

West London, as expected, put greater emphasis on leisure amenities, such as proximity to parks, proximity to good schools and older buildings, than the other submarkets. East London, however, put lesser values for the same leisure amenities. These results are logical as people living in higher house price neighbourhoods are generally willing to pay greater sums to live in proximity to leisure amenity such as open space (Anderson and West 2006). These results can be explained by the differences in attribute demand for different socio-economic grouping in London (Adair et al. 2000; McMillen 2012). West London, East London, the inner extension and the outer extension put more emphasis on living near active land uses such as shops. The results are also understandable as these submarkets are less car dependent and more reliant on public transport where proximity to shops matter more.

These results confirmed the existence of the London housing submarkets and clearly exhibited the different implicit prices and demand across the submarkets. These results also show using a general housing market average might substantially overestimate or underestimate the attribute value in particular housing segments.

7.5.4 Multilevel Hedonic Price Model Results

Results from the previous section showed St-HSs formed by spatial-structural-network factors passed the St-HS tests (Schnare and Struyk 1976), where implicit prices differed significantly between individual submarkets. The next stage in the analysis was to study the extent to which the St-HS effect existed when correlated with house prices through a multilevel hedonic price model. Table 7.8 reports the regression results for the first six candidate models.

Table 7.8 Multilevel hedonic price regression model results.

Street-based Housing						
Submarket	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a	Model 6a
Fixed						
Integration			0.046*** (0.008)	0.058*** (0.013)	0.035*** (0.012)	0.027** (0.012)
Floor Size			0.342*** (0.003)	0.340*** (0.003)	0.341*** (0.003)	0.277*** (0.003)
Age						0.030*** (0.003)
Park						-0.100*** (0.014)
Shops						0.031*** (0.005)
Flat						-0.093*** (0.004)
House						0.030*** (0.003)
_cons	12.660*** (0.006)	12.620*** (0.022)	12.620*** (0.020)	12.610*** (0.020)	12.630*** (0.089)	12.640*** (0.069)
Random						
var(Residual)	0.220 (0.004)	0.151 (0.003)	0.043 (0.001)	0.041 (0.001)	0.041 (0.001)	0.034 (0.001)
St-LA						
var(cons)		0.085 (0.010)	0.078 (0.009)	0.070 (0.009)	0.042 (0.005)	0.043 (0.005)
var(integration)				0.016 (0.003)	0.012 (0.002)	0.012 (0.002)
St-HS						
var(cons)					0.0528 (0.0298)	0.0313 (0.0185)
N	5,334	5,334	5,334	5,334	5,334	5,334
Loglikelihood	-3,536	-2,781	477	542	585	1049
LR test						
Prob > chi2	0***	0***	0***	0***	0***	0***

*** p<0.01, ** p<0.05, *p<0.

Model 1 was the null model, in which the grand mean of the log house price was 12.663; the residual illustrated that the total variance away from the mean was 0.220. This base model was discussed in the previous chapter. Model 2 was the level two varying intercept model, where the between-St-LA (level two) variance in house prices was 0.085 and the between property (level one) variance was 0.151. Model 3 was the level two varying intercept model with fixed predictors, where the between-St-LA (level two) variance in house prices was 0.078 and the between property variance was 0.043. The reduction

in the property variance was expected due to the inclusion of the predictors. Model 4 was the level two varying intercept and slope model with fixed predictors, where the between-St-LA (level two) variance in house prices was 0.070, the between-St-LA integration (level two) variance was 0.016 and the between property (level one) variance in house prices was 0.041. Model 5 was the level three varying intercept and slope model with fixed predictors, where the between-St-LA (level two) variance in house prices was 0.042, the between-St-LA integration (level two) variance was 0.012, the between-submarkets (level three) variance in house prices was 0.053 and the property (level one) variance in house prices was 0.041. Model 6 was a level three varying intercept and slope model with a wider set of fixed predictors, where the between-St-LA (level two) variance in house prices was 0.043, the between-St-LA integration (level two) variance was 0.012, the between-submarkets (level three) variance in house prices was 0.0313 and the within-property (level one) variance in house prices was 0.034. This reduction in the property variance was, again, to be expected, due to a wider set of fixed predictors.

The key findings showed significance for all six candidate models through the LR tests. The evidence suggested that both the local area effect and the housing submarket effect were significant. On the one hand, the local area effect remained relatively stable, with a small reduction due to the inclusion of fixed effect predictors and housing submarket variables. On the other hand, the submarket effect was reduced when the wider set of predictors were included. This result showed that the submarket effect overlaps with the predictor effect. This research found 4-5% of house price variations could be explained by the housing submarket effect when controlling for the local area effect and other predictor effects.

7.5.5 Street-based Housing Submarket Model Comparison

The chart below summarises the goodness of fit as measured by the AIC across the six candidate hedonic price models. The reduction in the AIC showed clear improvements in the statistical significance, which allowed the progressive inclusion of the local area effect, the submarket effect and the fixed predictor effect. These improvements showed that the property level effect, the local area level effect and the housing submarket level effect were all significant when correlated with house prices.

Table 7.9 Street-based housing submarket goodness of fit comparison.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
AIC	7,075.495	5,567.136	-944.593	-1,072.62	-1,156.56	-2,073.99

Figure 7.9 below summarises the ICC for each model. The ICC measured the amount of variation captured by the local area effect, the housing submarket effect and the property effect in proportion to the overall house price variance. Blue denotes the property variance, orange signifies the local area variance and grey designates the submarket variance. In Model 3, the property variance dropped to approximately 30% due to the inclusion of the predictors. The percentage did not drop below 30% of the overall house price variance for St-HSs. The variance was relatively similar across the three levels, after accounting for the predictor effects. This result, again, showed the nested effect of St-LA and St-

HS on house prices in London.

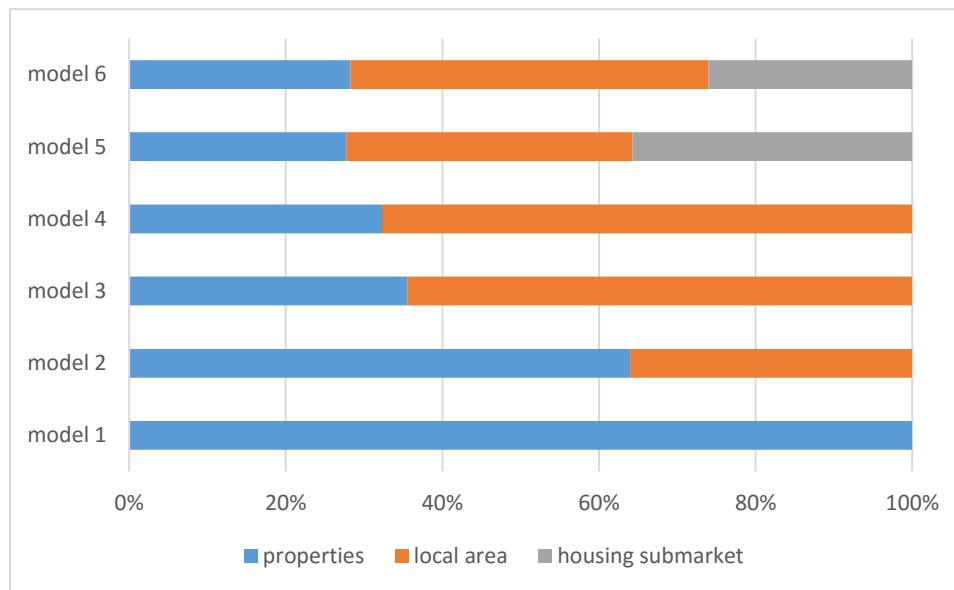


Figure 7.9 Comparison between the St-HS and the ICC.

For empirical reasons, spatial autocorrelation effects were also examined using the Moran's I index ($rad = 2400m$). The results did not differ significantly when adjusted for different radii. The global spatial autocorrelation was reduced from 0.27 in Model 1 to 0.0004 in Model 6. The probability still showed weak significance at the $p\text{-value} < 0.01$ level. These results confirmed previous studies, which argued that the inclusion of the neighbourhood and submarket effects in multilevel hedonic price models can also reduce spatial autocorrelation effects (Orford, 1999).

7.5.6 Street-based Housing Submarket and Traditional Housing Submarket Model Comparison

This section compares St-HSs and traditional housing submarkets using different local area units. Similarly, the LR test for the different housing submarkets was significant at the $p\text{-value} < 0.01$ level. For comparison, different housing submarkets were visualised. Figure 7.10 shows clear similarities between the five local area units. This was not surprising, as the same spatial-structural variables were used to construct the housing submarkets. However, despite the similarities, the differences in the results between the local areas were evident. An initial observation was that the smaller administrative local area units produced more fragmented housing submarkets. Second, the inner areas were much more clustered than the outer areas. Third, the St-HSs looked more clustered and less fragmented in general than the administrative local area housing submarkets.

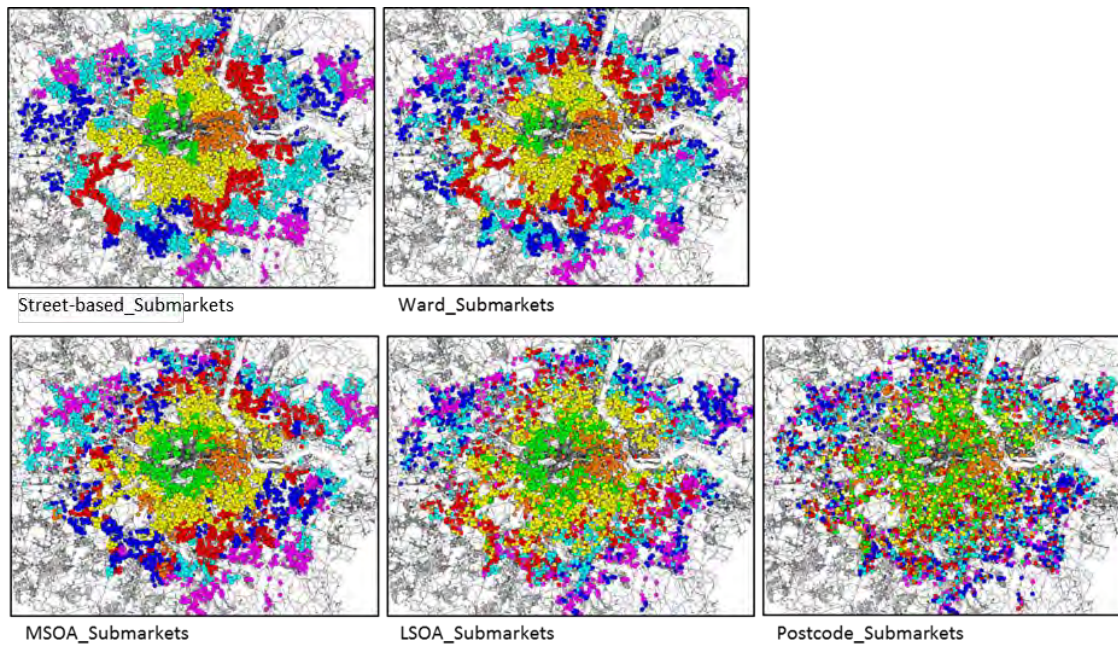


Figure 7.10 Visualisation of the St-LA, ward, MSOA, LSOA and postcode housing submarkets. Colours denote different housing submarkets.

This section estimated a multilevel hedonic price model to compare street-based housing and traditional housing submarkets. Figure 7.11 shows the goodness of fit, comparing the housing submarkets generated by the five local area units. Street-based housing submarkets (St-HSs) are shown in light blue, ward housing submarkets are coloured orange, LSOA housing submarkets are in grey, MSOA housing submarkets are shown in yellow, and postcode unit housing submarkets are coloured dark blue. The progressive, downward trend of the AIC showed that the three-level nested effects on the property were mutually significant. This evidence confirmed the hierarchical effects on house prices at the property level, the local area level and the housing submarket level. Importantly, this research also showed that housing submarkets formed by St-LAs achieved a lower AIC than housing submarkets formed by traditional region-based local area units. These results indicated that St-HSs had a stronger statistical fit than traditional housing submarkets when associated with house prices. These results also exhibited that the more clustered housing submarkets, such as St-LAs, wards and MSOAs, achieved a better result than the more fractured housing submarkets, such as LSOAs and postcode units.

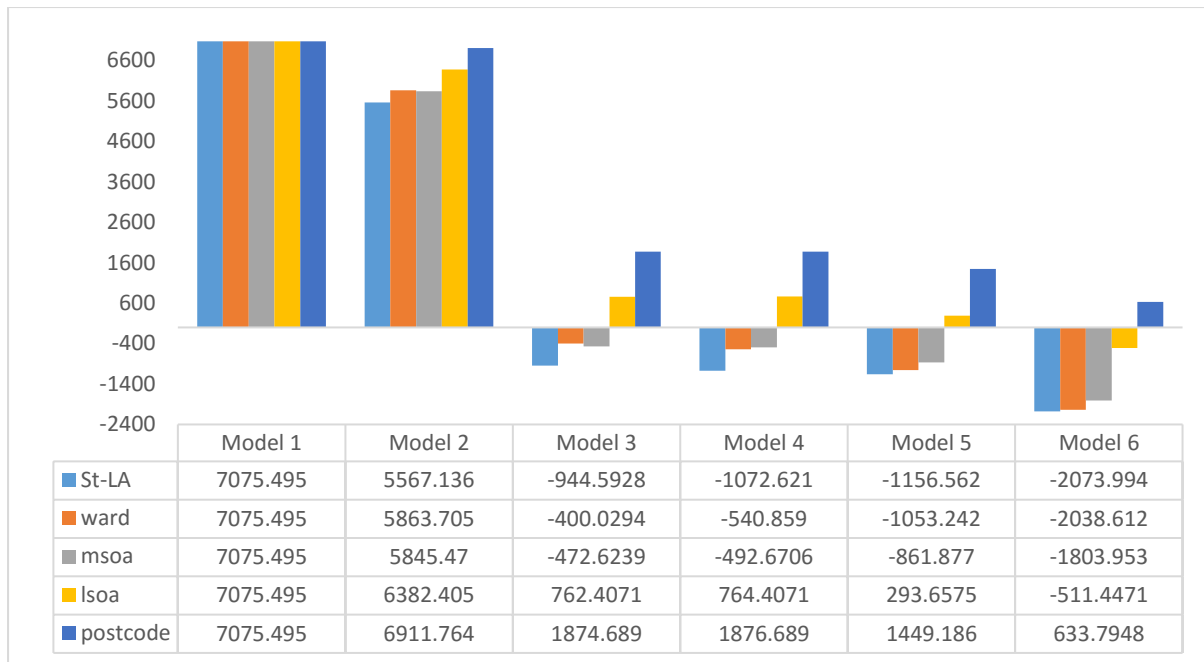


Figure 7.11 The AIC comparison across the different housing submarket models.

7.6 Discussion

The results from the submarket analysis confirm the existence of St-HSs in London, where seven submarkets with differing implicit prices are found. These seven housing submarkets are interpreted as west London, east London, the inner extension, the outer extension, the suburbs, the working class suburbs and edge London. The submarket definition seems to comply with the general understanding of the city, where the higher house price locations, such as West London, put greater value on school quality, access to shops and leisure activities, and lower house price locations in East London put lesser values on the same leisure amenities. Buyers outside of Central London generally have more bedrooms and a larger home, implying that this submarket buys more space to compensate for the losses in accessibility and amenities. This result further implies that spatial and structural features are inherently linked and that buyers in different submarkets simultaneously trade-off between spaces, amenities and accessibility in different quantities. One limitation to the submarket specification is the lack of differentiation between the North and the South, which might be due to omitted variables, such as the public transport variable, in the submarket construction.

The regression result shows significant nested property-local area-submarket effects on house prices in London (Orford, 1999). This research finds that 4-5% of house price variations can be explained by the housing submarket effect when controlling for the local area effect and other predictor effects. The research also finds that St-HSs have a stronger statistical fit and effect on house prices than traditional ones formed by administrative local areas. One interpretation is that housing submarkets shaped by St-LAs can capture more subtle differences in the urban environment than submarkets formed by ad-hoc administrative boundaries, since humans experience the urban environment at the street level. As a

result, the St-LAs might capture the perceptual dimension of housing submarkets more accurately, and thereby have greater effects on house prices in London. Administrative local areas, however, might divide the submarkets less accurately than St-LAs, thereby failing to capture the perceptual dimension of the housing submarket as accurately. The inaccuracy can lead to less homogeneity within the submarket and, as a result, a weaker effect on house prices in London.

The regression results also show that the LSOAs and the postcode units, which generated more fractured housing submarkets, generally have weaker effects on house prices than MSOAs, wards or St-LAs, which are more clustered. These results imply a strong clustering effects of the housing submarkets in London. The results also suggest that spatial configuration might be an important aspect in housing submarket formation. One plausible explanation for the spatial clustering effect is that buyers whom are attracted to live in a certain housing submarket would attempt to buy in a local area within the submarket or, if they cannot afford it, at the local area connected to it. This explanation might illuminate how the spatial clustered housing submarket definition has a stronger effect on house prices than the fractured housing submarket definition. This process also implies that housing submarkets in London are constantly shifting, with buyers in one housing submarket attempting to buy within the same submarket or, if they cannot afford it, near the submarket's edges. For example, those who want to buy in Shoreditch have shifted to Dalston. Those who want to buy in Dalston have shifted to Hackney. This shift can be seen in the house price per square metre difference between 1995 and 2011, as shown in figure 7.12 below. To prove this process, further empirical work needs to be conducted in order to see how the housing submarket boundaries have shifted. The concluding chapter will discuss some dynamic techniques, such as the use of cellular automata in modelling the housing submarket, to explore this process.

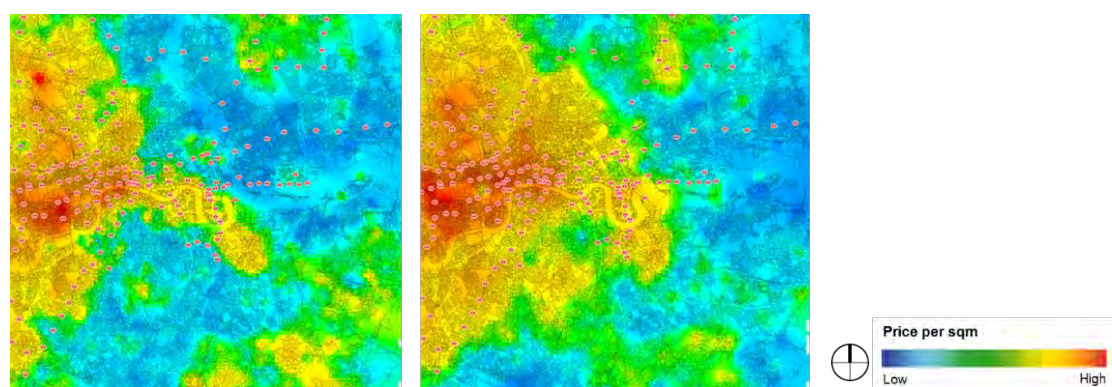


Figure 7.12 London house prices per square metre for 1995 (left) and 2011 (right).

7.6.1 Conclusion and Limitations

The benefits of using St-HSs are two-fold. First, St-HSs consider the spatial configuration and the perceptual dimension. Second, St-HSs improve the statistical fit of the hedonic price model, thereby giving more accurate results empirically. The results have also enhanced our understanding of housing submarkets and how spatial configuration affects them. This finding can lead to greater efficiency in

housing policies, improve the supply of housing services and potentially predict where new housing submarkets will shift to in the future.

There are limitations to the use of St-LAs in defining housing submarkets. Similar to the previous research strand, St-LAs only considered the topology of the street network. As a result, this method did not consider other aspects of the urban environment, such as the massing or the architectural form, in differentiating local areas. Taking into account other urban environment qualities, such as architecture and streetscape, is an obvious way of extending this research. There are also noticeable socio-demographic factors excluded from the submarket construction, such as income (Schnare and Struyk 1976; Strazheim, 1974) and ethnicity (Palm, 1978). This is important as there is extensive research studying the heterogeneity of attributes demand across different socio-economic housing submarkets. Theriault et al. (2005) for example have found that the effects of shops on house prices can differ according to both the types of shops and the social demographics of those living near them. Similarly, Anderson and West (2006) have found that the effects of parks on house price can differ according to the income, location and density of an urban environment.

Methodologically, the k-means clustering method has many limitations, such as the a priori-determined cluster number and the use of a simple cost function (the sum of the squared error) to measure the cluster quality. As the objective of the research was to compare different local area units, the use of a standard unsupervised clustering method was justified. However, more complex, non-linear classification methods, such as the use of machine learning methods (Kauko 2002), could be considered, as they could potentially provide more accurate results. With the advent of large-scale behavioural data, recent research has also defined housing submarkets using this behavioural data. For example, a study from Park (2013) demonstrated households in a similar housing submarket exhibit similar travel patterns. Recent studies have also seen the use of fuzzy logic methods to identify overlapping housing submarkets (Helbich, 2015), where submarket definitions were found to shift over space and time (Jones et al. 2003). Future research should consider studying the dynamics of spatial housing submarkets over space and time.

To summarise, housing submarkets are complex in formation and definition. They are determined by multiple factors that overlap and constantly shift across space and time (Watkins, 2001). This research demonstrates that similarities across properties in the same submarket extend beyond sharing similar structural and location attributes to include spatial network attributes. This research reveals that the configuration of the spatial network is simultaneously important at the property level, the local area level and the submarket level. This result is not only important for real estate economists in the valuation of housing but also in improving the housing submarket definition for informing housing policies. Importantly, this perspective illustrates the importance of understanding the housing market as a complex system of connected properties at the street level, the local area level and the housing submarket level.

Chapter 8

Conclusion

8.1 Introduction

This chapter will summarise the thesis, identify its implications, limitations and future research directions. It is organised as follows. Section 1 introduces the chapter. Section 2 outlines the thesis chapters and summarises the thesis research findings. Section 3 identifies the key research limitations. Section 4 presents future research directions. Section 5 describes various research implications, and Section 6 concludes the thesis. The following section will discuss the key findings for each research strand.

8.1.1 Chapter summary

In Chapter 1, this research began by asking, 'What is the effect of spatial network configuration on house prices?' This question is important, as a better understanding can help advance urban design through improved evaluation methods, more efficient financing and incentives for developers to design better places. In order to answer the research question, Chapter 2 introduced the hedonic price approach as a set of theories and statistical techniques, including space syntax methods, to capture the economic value of urban form. Chapter 3 presented the three spatial configuration methods used in the empirical analysis, which consisted of utilising the spatial network to measure the geometric accessibility, street-based local areas and street-based housing submarkets. Chapter 4 discussed the analytical chapters, the Greater London case study and the various datasets. Chapters 5, 6 and 7 were the key analytical chapters. Chapter 5 concerned the estimation of the geometric accessibility effect on house prices. Chapter 6 considered the assessment of the street-based local area effect on house prices and Chapter 7 included the valuation of the street-based housing submarket effect on house prices. All three analytical chapters, discussed in detail in the next section, found spatial configuration has a significant effect on house prices.

8.2 Key Findings

Bringing it all together, spatial configuration effects were significant across all three analytical chapters. Figure 8.1 illustrates the intra-city hedonic price model framework where house price can be decomposed into its constituents such as location attributes, structural attributes, neighbourhood attributes, and the housing submarket it sits in. This research adds to this framework by adding geometric accessibility effects, street-based local area effects and street-based housing submarket effects into the intra-city hedonic price model. These results reveal, there are significant economic value within the spatial configuration of cities in structuring these spatial goods. The implications are great as retrieving the relative economic value of spatial configuration across these three levels could bring greater efficiency in allocation resources leading to more optimal design of cities and neighbourhoods (Webster 2015). This section will describe in more details the key findings for each research strand. Some figures are reproduced for this chapter to facilitate the discussion.

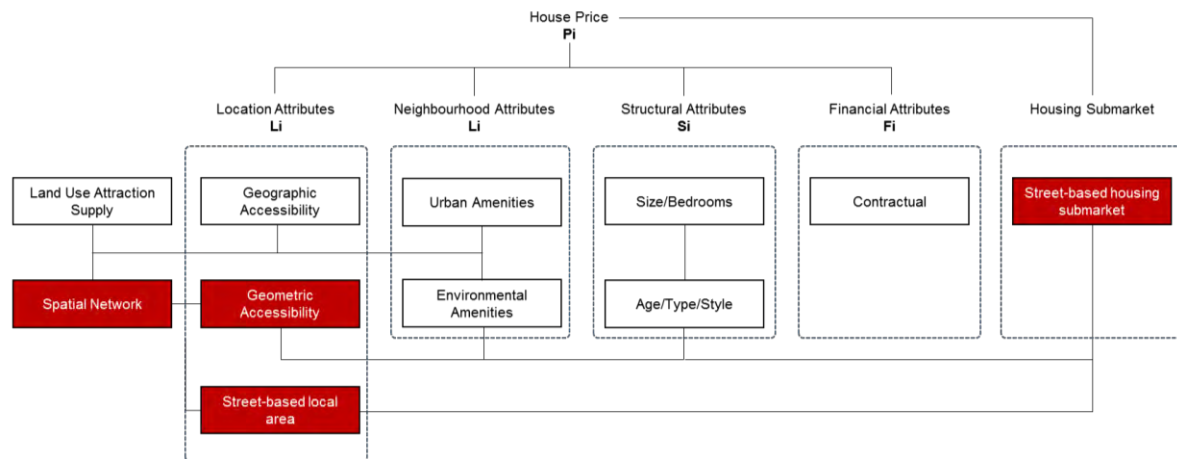


Figure 8.1 Intra-city Hedonic Price Model

8.2.1 Spatial Network Accessibility

Research Strand One question: To what extent are measures of geometric accessibility associated with intra-city house price variations? how does geometric accessibility compare with geographic accessibility measures when associating with house price?

To answer the first question, this research examined the correlation of geometric accessibility measures (closeness centrality) and geographic accessibility measures (gravity-based accessibility) with house prices through a hedonic price regression model. The first key finding of the research provides evidence that confirms the effect of geographic and geometric accessibility improvements (e.g. key transport projects in London between 1995 and 2011) on house prices. This significance is to be expected, as the transport projects increased access to employment, amenities and social opportunities. The research also finds a strong overlap between the two accessibility measures, implying places with the greatest connectivity and social opportunities also have the highest number of employment opportunities as highlighted in figure 8.2. One explanation is to relate back to Hillier's (1996) theory of the movement economy, where the grid is not only the generator of "natural movement" but also the generator of "employment activity" and thus "destination movement." This coupling between spatial configuration and employment suggests geometric accessibility is also capturing employment accessibility effects. This result also suggests geometric accessibility can be used as an "accessibility" measure in hedonic price models when data for geographic measures is not available. This measure can be especially useful in the UK, where official Census data is only gathered in an aggregated format and published every ten years.

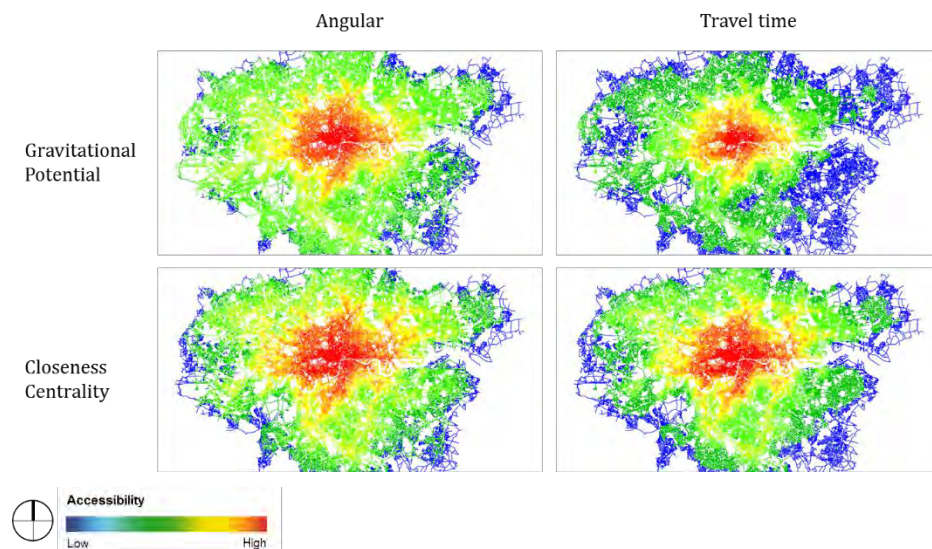


Figure 8.2 Space Syntax Integration (Closeness Centrality) and Gravitational Potential. Accessibility measures mapped by a colour spectrum, where red denotes high accessibility and blue denotes low accessibility.

The second key finding of the research reveals that the two accessibility measures have subtle, differential effects on house prices. There are several plausible reasons why these two measures differ in how they explain house prices. One explanation is that, when purchasing a property, a buyer is not only acquiring access to employment, shops, parks and schools, but also opportunities related to spatial connectivity. This can come from the greater social opportunities afforded by the denser grid in the city centre (Jacobs, 1961) or simply the general attraction effect that geometric accessibility captures (Webster, 2010).

The third key finding of the research shows that house prices correlate positively with spatial network closeness centrality and negatively with spatial network betweenness centrality. This effect is logical, as buyers do not simply purchase accessibility to central places, but also prefer protection from the noise and pollution of high betweenness spaces. These results suggest a complex trade-off on being central (closeness centrality) at the city-scale but also being isolated locally, being proximate to amenities but being one step away from the main routes (betweenness centrality). However this effect also differs between different main street. Figure 8.3 shows in detail that house price is lower on the main street along Green Lanes.

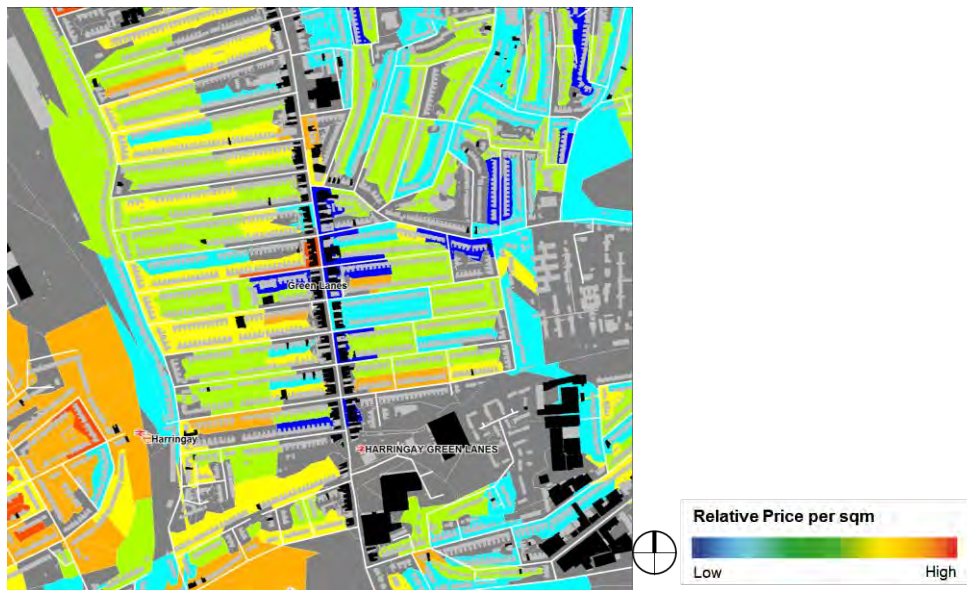


Figure 8.3 Green Lanes normalised (1995-2011) house price per sqm.

The fourth key finding of the research is that the accessibility measures for radius 60mins are preferred to those for radius infinity. The results are logical, as the average commuting time in South East England is between 50-70 minutes. Having greater opportunities to access beyond the average commuting time will logically offer less marginal benefits to the users. To conclude, the results confirm and extend the established relationship between geometric measures and house prices in London. A key contribution of the analytical chapter is that the use of spatial configuration techniques can provide a deeper understanding of accessibility effects on house prices. To generalise the research results, further investigation across different geographical regions, different specification and time periods are necessary.

8.2.2 Street-Based Local Areas

Research Strand Two question: To what extent do St-LAs, as defined by the topology of the street network, associate with house prices? Secondly, how do St-LAs compare with administrative local area units when correlating with house prices?

To answer the second research question, this analytical chapter applied community detection techniques to the street network dual graph in defining Street-Based Local Areas (St-LAs) in London as shown in figure 8.4. The St-LAs were then compared to the correlation between the administrative region-based local area and house prices through a hedonic price regression model. The first key finding is that house prices are more similar within St-LAs and that house price variations between local areas are greater when the local areas had fewer connections. The second key finding from the regression results shows local areas have significant effects on house prices and that St-LAs are preferred to administrative units. There are three plausible reasons for the house price effects.



Figure 8.4 London Street-based Local Area in identifying the Isle of Dog. Colours denote different St-LA.

Firstly, people experience the urban local area along a street network. The street network is, therefore, able to more precisely capture subtle differences in an urban environment and more accurately the perceptual definition of a local area than an ad-hoc administrative region. The probability of walking within the same street network community is greater than in another subgraph. This greater probability implies that there might be a linkage between spatial network clusters and the collective perception of neighbourhoods, as afforded by the spatial configuration of the street network. To verify this, further empirical research comparing individual perception maps to street-based local area units is needed.

Secondly, these street network communities reinforce socio-economic similarities within the local area, as like-minded people cluster together and bump into each other. This clustering emphasises the socio-economic similarity within and the boundaries between St-LAs. Plausible processes allowing this to happen include crowd behaviour and bounded rationality (Benerjee, 1992; Simon, 1957). Figure 8.5 from chapter six offers an argument in how spatial configuration can influence street-based-local-area formation through a simulation. The figure illustrates two simulations of an agent who starts from different orange nodes and randomly walks around the graph. The first simulation shows the walker took nine steps to reach the bridge. The second simulation shows the walker took eight steps to reach the bridge. This demonstrates that a random walker is likely to stay longer within a local area when there is greater intra-cluster connectivity.

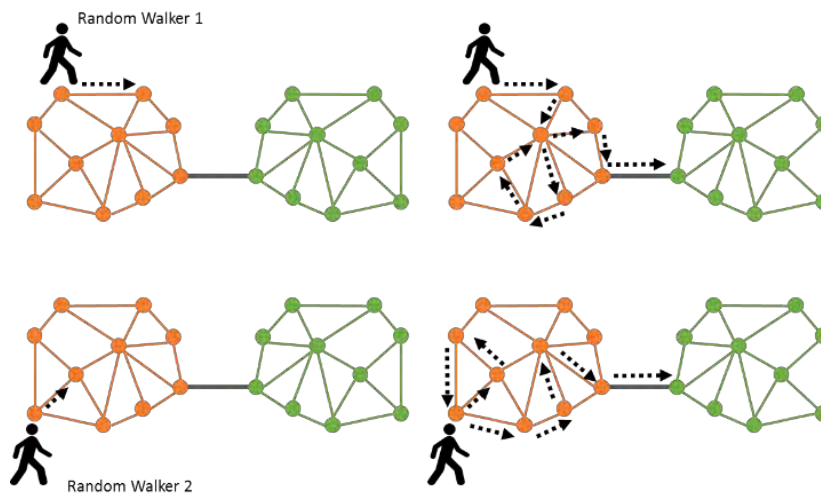


Figure 8.5 Random walker simulation.

Thirdly, local areas can become part of the housing bundle. For example, when a buyer purchases a property in Kensington or Crouch Hill, they are also buying a Kensington or Crouch Hill local area premium as part of the housing bundle. Therefore, a buyer values the house similarly to another within the same area, rather than to a house in a different area. This effect is related to Tobler's first Law of Geography, where properties that are closer to each other are likely to be more socio-economically similar than properties that are further apart (Tobler, 1970). Over time, this implies a further reinforcing effect on house prices. To sum up, the main contribution of this research strand is the novel application of community detection techniques on the street network dual graph that defines St-LAs in London. The results show that local areas have a significant effect on house prices, and that St-LAs have a stronger statistical fit than traditional region-based local area units.

8.2.3 Street-Based Housing Submarket

Research Strand Three question: To what extent are Street-Based Housing Submarkets comparable to traditional Census-tract based housing submarkets when associated with house prices?

To answer the third research question, the analytical research applied statistical clustering to identify Street-Based Housing Submarkets (St-HS). St-HSs were then compared to administrative region-defined traditional housing submarkets in correlating with house prices through a hedonic price regression model. The main contribution of this research strand is the unique application of street network configurations in defining St-HSs for London. This research finds that St-HSs defined by St-LAs are significant and are preferred to housing submarkets defined by region-based administrative units.

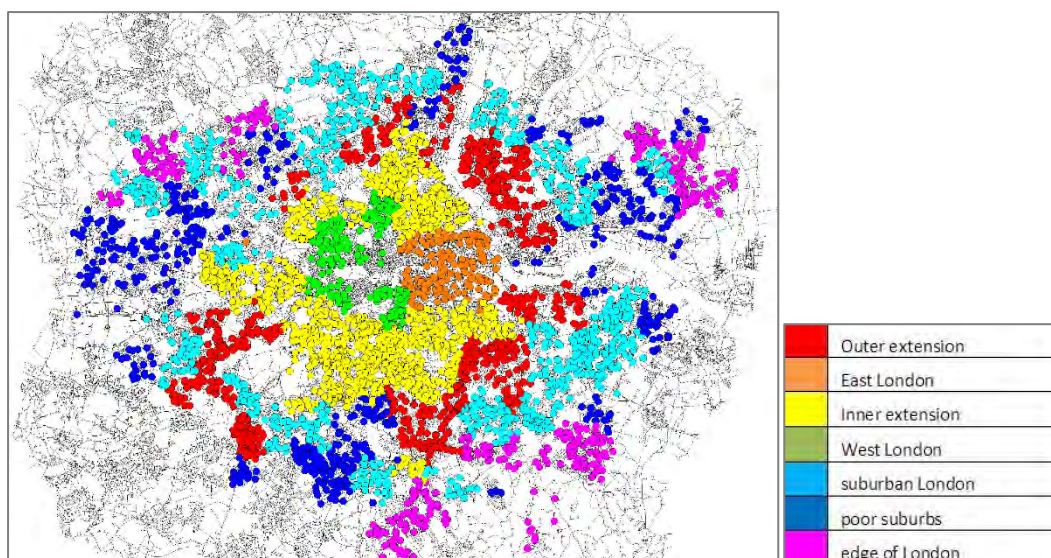


Figure 8.6 Greater London St-HSs. Colours denote different housing submarkets.

The first finding is that the submarket analysis confirms the existence of St-HSs in London, where seven submarkets are found as shown in figure 8.6. These seven housing submarkets are interpreted as West London, East London, Inner Extension, Outer Extension, Suburbs, Working Class Suburbs and Edge London. The submarket definition complies with the general understanding of the city, where the pricier house locations, such as West London, place greater value on school quality, access to shops and leisure activities; cheaper house price locations, such as East London, place lower value on the same leisure amenities. These results are logical as people living in higher house price neighbourhoods are generally willing to pay greater sums to live in proximity to leisure amenity (Anderson and West 2006). These differences in demand can potentially be attributed to locational and demographic differences where different spatial housing submarkets are making trade-offs between affordances of space, amenities and accessibility (Adair et al. 2000; Anderson and West 2006; Theriault et al 2005). These results also point to the underlying problem in using an average housing market attribute value could substantially overestimate or underestimate the attribute value in particular housing segments.

The second finding is that a housing submarket formed by an St-LA has a stronger effect on house price than traditional housing submarkets formed by local administrative areas. One interpretation is that St-HSs capture subtle perceptual differences in the urban environment better than traditional housing submarkets formed by ad-hoc administrative regions. Administrative local areas, might be dividing the submarkets less accurately than St-LAs, thereby leading to less homogeneity within the submarket and, as a result, a weaker effect on house prices in London.

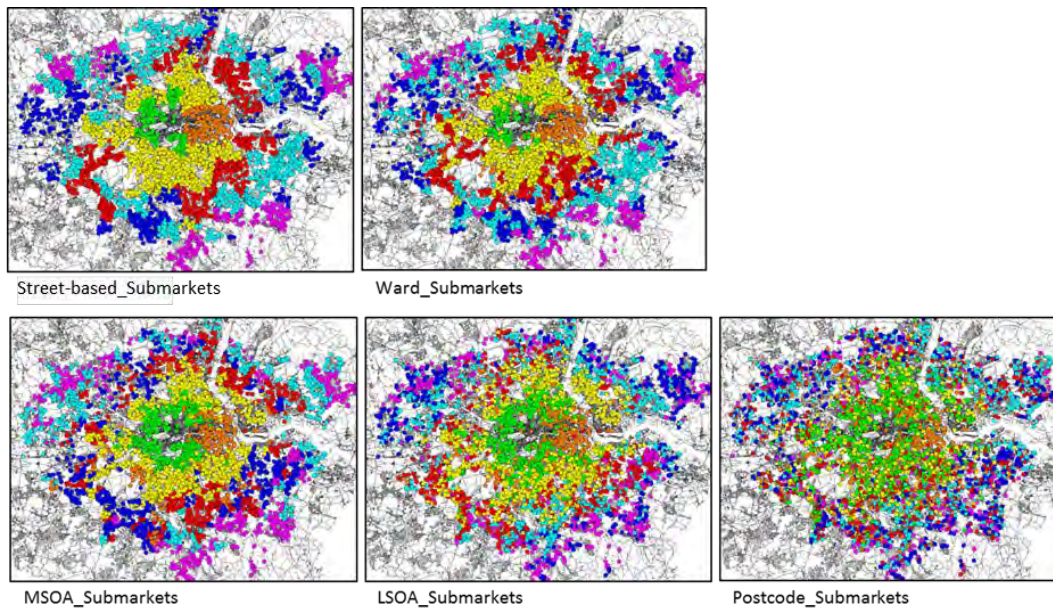


Figure 8.7 Visualisation of the St-LA, ward, MSOA, LSOA and postcode housing submarkets. Colours denote different housing submarkets.

The third finding is that LSOAs and postcode units, which generate more fractured housing submarkets, generally have weaker effects on house prices than St-LAs, MSOAs or wards, which are more clustered as shown in figure 8.7. These results imply a strong clustering effects of the housing submarkets in London. One plausible explanation for the spatial clustering effect is that buyers whom are attracted to live in a certain housing submarket would attempt to buy in a local area within the submarket or, if they cannot afford it, at the local area near the submarket's edges. For example, those who want to buy in Shoreditch but cannot afford to might prefer to buy in Dalston and those who want to buy in Dalston and cannot afford might prefer to buy in Hackney. This can be seen in figure 8.8 where we see the shifting ridgeline of house price in London. Housing submarkets dynamics have not been studied in this research but will be briefly discussed in the limitation and future works.

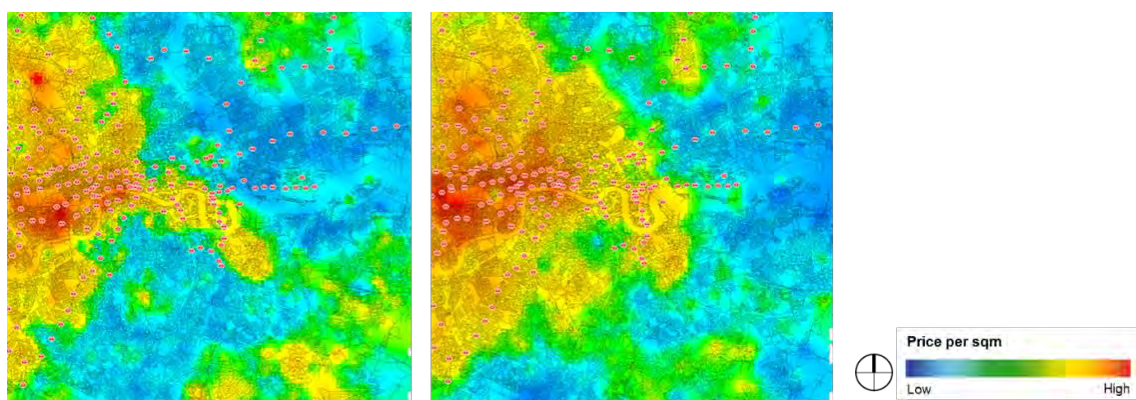


Figure 8.8 London house prices per square metre for 1995 (left) and 2011 (right).

8.2.4 Summary and Potential Causal Inferences

To summarise, a number of conclusions can be drawn from this research. First of all, this research empirically demonstrates that spatial configuration, or the geometrical properties of the built environment, significantly affects house prices and the economic outcome of a city. This effect is observed at the property level via accessibility, at the neighbourhood level via the local area and at the district level via housing submarkets.

A key question has not yet been discussed: What does all of this information mean and what is the causal mechanism via which this process takes place? As correlation does not imply causation, these early results are not able to infer causality, though they do point in useful directions where this causal mechanism could take place. One research direction is that spatial configuration influences residential location choices and thereby the house prices where the market operates.

This inference can be drawn from the theory of natural movement and movement economy (Hillier et al. 1999), where space syntax theory provides a strong argument that spatial configuration is the prime cause for natural movement patterns and thus, in turn, drives a market for commercial uses, which acts as the attractor for the destination movement pattern. Building on this notion, this research takes this argument one step further by suggesting that spatial configuration not only produces the spatial network accessibility effects of centrality and employment but also influences the construction of the community and the spatial clustering of the housing submarket. The spatial configuration of the built environment thus impacts residential location choices and prices across multiple scales.

For example, the choices individuals make concerning residential location are influenced by spatial configuration, which influences employment location. Linking to the bid rent theory, this process, in turn, influences the prices individuals are willing to take when they trade-off between transport costs and house prices. In addition, concerning the use of community detection in defining St-LAs, the choices that individuals make regarding residential location are influenced by spatial configuration, which influences community formation with which individuals associate. The way in which markets operate by pricing the value of that community for different individuals is a way that society spatialises this process. This effect is similar in the spatial clustering of housing submarket. The multi-scale effects of spatial configuration on house prices, as demonstrated in the three chapters, support this inference.

These initial conclusions, supported by the spatial configuration theory, offer insight and suggest a potential research direction, within which spatial configuration can influence residential location choice and house prices. Further empirical research is required to examine these inferences. In the next section, this thesis summarises the limitations of the research.

8.3 Research Limitation

The most apparent constraint is the attention on the single case study of Greater London. Focusing on a large metropolis, such as Greater London, is important, as recent population growth is mostly

seen in metropolitan regions. Examining a single case study also allows for a more in-depth look of the three scales: property level, local area level and submarket level. However, investigating a single case study diminishes the generalisability of the research. As seen from previous investigations, spatial network effects on house prices can differ significantly between geography, income and the size of cities (Law et al., 2017a). Further research across longer periods of time might also reveal differences in the value of accessibility, where there are significant differences in demographic and economic structure.

Secondly, there are several econometric limitations to the research. Despite the novelty in the use of a multi-level hedonic price model in capturing the spatial hierarchical effects (appendix C) on house price, there could be spatial effects within the local area that might not be uncovered. Further-more, there are also complex non-linear relationship between independent variables that has not been modelled. As the aim of the research is to establish the relationship between urban configuration and house price, complex specification in econometrics has not been adopted. There is a necessity for future research to consider using more robust methods such as spatial panel data model in capturing spatial temporal dynamics and artificial neural network model in capturing complex non-linear effects. As the research focus is architecture and urbanism, more complex spatial econometric methods have not been adopted.

thirdly, there are noticeably missing variables in the general hedonic price regression model, such as considering the effect of urban design factors (Nase et al. 2013), social accessibility and across different socio-demographic housing submarkets (Adair et al. 2000; McMillan 2012). These exclusions should be considered for future research in order to improve the model's overall predictive accuracy, reducing bias from omitted variables but more importantly to be able to identify new insights. For example, an obvious questions is, what is the effect of spatial configuration for different income groups?

Extending from the key research findings, there is also a need to better understand how spatial configuration affects house prices at an individual level. For example, how does the spatial network influence buyers and sellers during the property purchasing process? Recent research looking at how social accessibility impact on residential location choice is beginning to address this topic (Tounonen and Law, 2017). In the following section, the specific research strand limitations are presented.

8.3.1 Research Strand One Limitations

There are several benefits to using geometric accessibility measures. Firstly, these measures do not require employment location information in measuring the accessibility. Detailed employment distribution data is not easily available in the UK, where the Census data is only gathered every ten years. Secondly, geometric accessibility measures are able to capture general accessibility effects, as afforded by its spatial configuration (Webster 2010). However, this brings up a number of limitations to using these measures. One constraint is that it is unclear what general benefits geometric accessibility effects are capturing. For example, is the measure encapsulating the accessibility

benefits of the grid, the social opportunities that the grid provides or various unexplained factors not being observed?

A second limitation is that the study used a simplified spatial network model when computing the accessibility measures. The spatial network accessibility model currently fails to calculate separate accessibility measures between cars and pedestrians, fails to consider other modes of transport (e.g. the bus network), and fails to consider the frequency of the service.

A third limitation is that the study did not consider the actual travel behaviour of the residents. For example, improvements in accessibility do not necessarily transfer to commuting behaviour change. Hence, further research is required to examine not only more sophisticated spatial network models, but also to consider using commuting behaviour data when measuring location differences in hedonic price regression models. Further research is also needed to encompass inter-regional transport projects. For example, the future High Speed Two project will make inter-city commuting more desirable. Housing markets could begin to overlap and influence each other, such as London and Birmingham. All of these topics require additional research to better understand the accessibility effect on house prices.

8.3.2 Research Strand Two Limitations

There are a number of strengths and weaknesses to using St-LAs in housing research. Firstly, using the street network as the geographic unit might encompass the perceptual dimension, which has the potential to reduce the arbitrariness of local administrative area units. Secondly, as the street network is the most permanent of all morphological elements, the use of St-LAs also allows for consistent comparisons across time.

The use of St-LAs have similar limitations. The first is that only using the street connectivity matrix in defining neighbourhoods is contrary to the belief that neighbourhoods are made up of many overlapping factors (Lebel et al., 2007; Galster, 2001; Kearns and Parkinson, 2001; Alexander 1965). For example, a local area can be identified from its architectural style or the social demographics of the area but that this identity can also shift across space and time. A case in point is that some areas in London such as Brixton which were once seen as a working-class neighbourhood is now desirable to live. Similarly, this technique in recovering street communities should also be tested across geographical regions. These techniques might be less relevant in the context of regular grids with a high degree of uniformity such as the cities in America.

Secondly, this research only tested the most commonly used community detection techniques using qualitative ground-truth. Since the write-up of the thesis, several promising community detection algorithms have been proposed. This includes the use of percolation theory in identifying network clusters from a distance threshold which successfully recovered the different functional and political regions in the UK (Molineros et al, 2015). Another promising approach known as COMBO from MIT combines the three elementary actions (merging, splitting and moving) of community detections into a single algorithm (Sobolevsky et al. 2015). These techniques should be tested in future research.

8.3.3 Research Strand Three Limitations

The strength in using St-HSs is that, like St-LAs, spatial configuration attributes are considered in its definition. The key limitation in the use of St-HS is methodological. For example, the use of unsupervised clustering, such as K-means clustering, is justified in this context, but there are more sophisticated methods available. This includes the use of fuzzy logic in identifying overlapping housing submarkets and the use of machine learning classifiers find optimal housing submarkets (Helbich, 2015). Future research should also consider other variables for submarket identification. This could include the use of traditional housing submarket variable such as income (McMillen 2012). For example, Xiao et al. (2017), assessed the people's willingness to pay for an amenity such as green space across different housing segments. This could also include the use of behaviour data, such as commuting flow patterns, in identifying commuting housing submarkets (Park, 2014).

8.4 Future Research Directions

From these limitations, multiple research directions can be identified. This section describes four of these paths. The first aim concerns the need to examine the spatial network effects for different geographical regions. The second research direction concerns the necessity to better understand local areas and submarkets as multi-dimensional overlapping areas. The third research angle concerns the demand to better understand how spatial configuration processes affect house prices. The fourth research perspective concerns the need to better understand the economic value of local urban design.

8.4.1 Case Study Across Geographies

We recommend first to examine the spatial configuration effect on house prices across different geographical regions, sizes of cities and types of cities. This would highlight the generalisability and specificity of the research into how spatial network configuration influences house prices across the three levels, but to also highlight the extent this effect differs between cities, socio-economic-political environment and geographical context. For example, do vehicular-dependent cities or smaller cities value accessibility less? One could conjecture that car-dependent cities would value accessibility less because residents with cars would be much less affected by distance. Figure 8.9 below shows UK house prices on the right and spatial network closeness centrality on the left. The image shows some similarity at the regional level, where South-East England, with the highest centrality, also has the highest house prices.

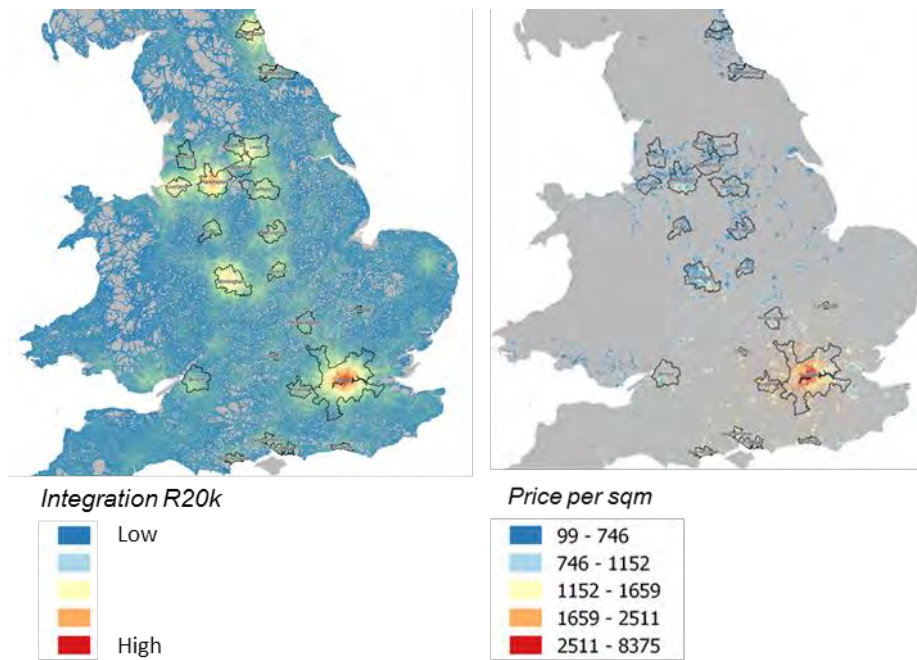


Figure 8.9 UK spatial network closeness centrality on the left and house price on the right.

However, Figure 8.10 illustrates that when we zoom into Birmingham, there are dissimilarities between house price and accessibility. Properties with the highest house prices are located near large country parks.

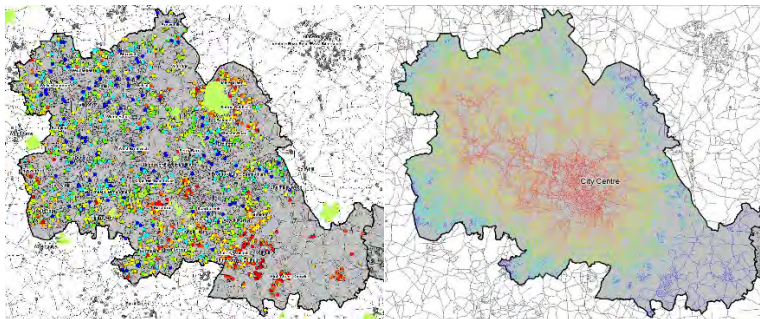


Figure 8.10 Birmingham spatial network closeness centrality on the left and house price on the right.

A general regression model included the house price and the spatial accessibility for London and Birmingham as an initial comparison.

$$\text{Log}P_i = \beta_1 \text{Int} + e_i$$

Where

LogP_i is the Log-price-per-square-meter of a property for i = postcode

Int represents the closeness centrality variable

B are the coefficients for the independent variables

e is the error term

Equation 8.1

Figure 8.11 below shows the regression result and scatterplot between London on the left and Birmingham on the right.

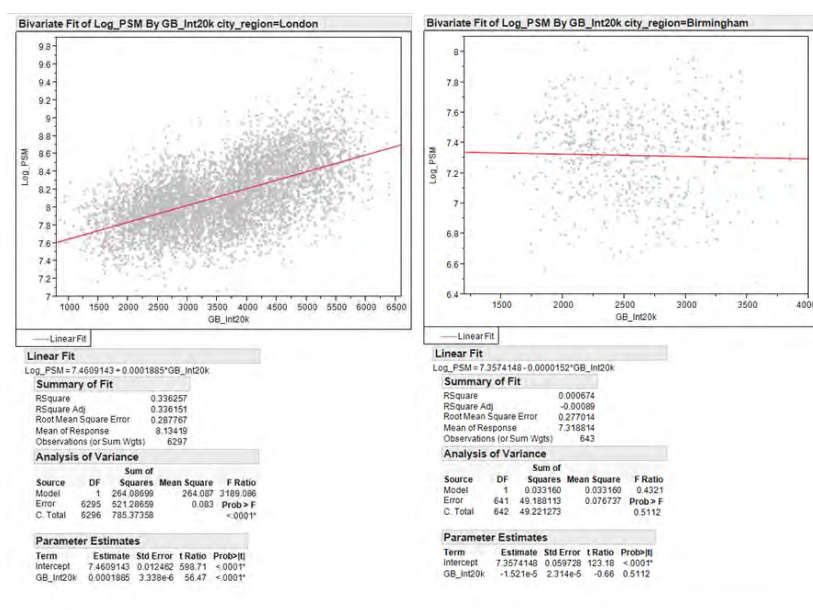


Figure 8.11 Scatterplots between house price per square metre and spatial network closeness centrality in London on the left and in Birmingham on the right

These results clearly show that the accessibility effects in the UK can differ significantly across different geographies. In some cities, such as London, the economic value of accessibility is significant, whilst in cities like Birmingham, it is not. Appendix D shows the coefficient differences across cities in the UK. A recent paper from the author (Law et al, 2017a) shows cities that value central places more tend to be denser, more productive, had greater proportion employed in the education sector and, most importantly, were less car dependent. This fits with emerging hedonic price research that illustrates hedonic price differentiates geographically.

8.4.2 Fuzzy Spatial Network Local Area Boundaries

Another key limitation of this research is that sharp and discrete boundaries have been created by community detection methods. Distinct boundaries are needed for practical reasons, such as in terms of land ownership and policy definition. However, this definition is also unrealistic at a perceptual level, as local areas are clearly multi-dimensional, overlapping and fuzzy, akin to a semi-lattice network (Alexander, 1965). This multiplicity is also reflected in the fact that there are multiple 'ground truths' in communities.

As a result, one future research direction is to define a fuzzy St-LA boundary. There are several methods to achieve this, including the application of fuzzy logic (overlapping memberships) in community detection. One can apply fuzzy logic to identify core and fuzzy areas. In this case, core areas are defined as spaces that are consistently in the same local area; fuzzy areas are spaces that

regularly overlap with different local areas. Figure 8.12 was generated by running the modularity optimisation algorithm multiple times; it demonstrates these types of core and unclear areas, where the green lines represent core areas and the red lines are fuzzy areas. The outcome of the algorithm changes with different runs. The result below shows that central areas in London, such as Soho, have a fuzzier definition due to their more porous street network. The next stage in the research is to understand the fuzzy local area effect on house prices. For example, to what extent do fuzzier and clearer local areas have greater or lesser homogeneity? Again, this fits with the emerging hedonic price research in identifying ambiguous housing submarket boundaries (Helbich, 2015).

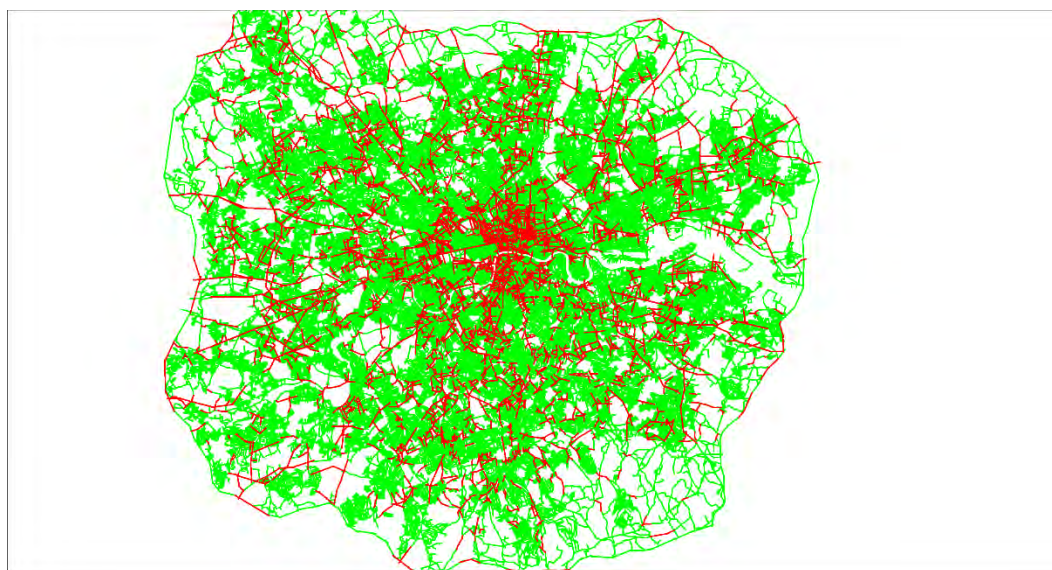


Figure 8.12 Local area cores (green) and fuzzy boundaries (red).

8.4.3 Simulating House Price Spillover Effects

At the heart of this thesis, the effect of spatial configuration on house prices was modelled using econometric methods. A criticism of these methods is that association does not necessarily lead to causation. Despite consistent evidence of connections between spatial configuration and house prices, there is a lack of research into how configuration influences prices. There may be other hidden effects that statistical methods have not uncovered. One approach is to adopt a mixed-methods approach, where both qualitative and quantitative methods are used in the analytical research. This includes examining the residential location decisions through structured interviews. For example, how does spatial configuration influence the residential location choice and the actual property bid process? This understanding of how social accessibility affects residential choice location is demonstrated by recent research by Tounonen and Law (2017).

A second approach is to use a complex-system approach, such as cellular automata or agent-based models, to understand dynamic effects within the system. The most famous of these is the Schelling (1967) model, which uses simple agent behaviour to model the emergent behaviour of social segregation. These methods, such as the agent-based model, have previously been applied in space

syntax to model pedestrian behaviour (Turner and Penn, 2001). However, there is no existing research that considers using such methods in modelling the spatial configuration effect of a real estate transaction. This reluctance stems partly from the lack of data in monitoring the micro-temporal, dynamic process of housing transactions, and more importantly from the lack of interdisciplinary research between spatial configuration research and complex system modelling.

For example, the use of the cellular automata modelling framework can be adopted to study the diffusion of house prices. A key research gap is the predominant use of cell-based methods rather than street-based methods in modelling the diffusion process. By adapting the cellular automata model for use in a dual graph, this research can bring together the foundation of spatial configuration methods with complex system modelling. These methods would allow the examination of spatial network effects on house prices over time. Figure 8.13 shows how dual-graph automata models can be used to examine the influence of the spatial network on house price diffusion, using Greater London as a case study. This abstract model assumes the street network is classified according to three house price categories. Furthermore, the model uses data from 1995, where blue shows low house prices, yellow represents medium house prices and red designates high house prices. The objective is to observe whether these simulations are able to replicate future house price distributions. The stochastic simulation starts where house prices in one segment change probabilistically according to the house price levels of the surrounding segments. Figure 8.13a shows the first step of the simulation, and Figure 8.13b shows the 20th step. Differences between the two images show how the spatial configuration of the street network can influence the speed of house price diffusion. Future research using observed house price data will be conducted to validate the model.

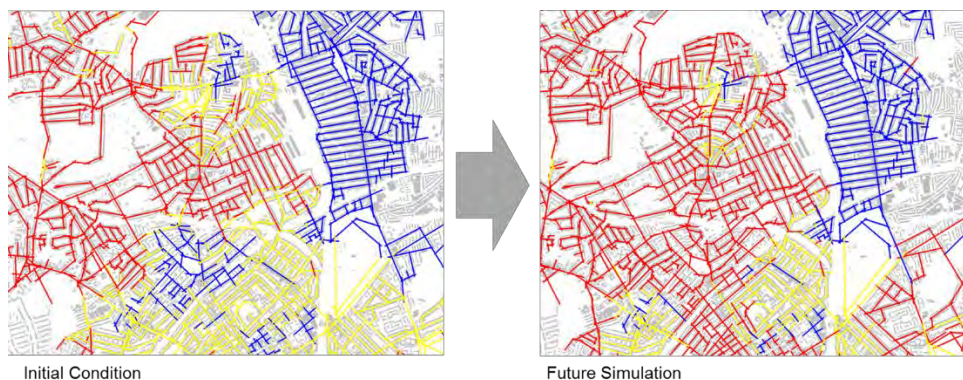


Figure 8.13 House price spillover simulations.

- a. House price simulation in step one
- b. House price simulation in step twenty

8.4.4 Estimating the Economic Value of the Urban Design

Last, there is also a need to examine the economic value of architectural and urban design, which impact people's social, economic and health outcomes. Nase et al. (2013) have begun to examine such effects and have found that urban design features such as façade continuity, variety, massing

and materiality have had significant effects on house prices when using the hedonic price approach. Limited research have study this problem which can be attributed to the cost and time necessary to collect this data. One approach to study this problem is to cast it as a machine vision classification problem. Law et al. (Forthcoming) have used a deep learning approach (Jia, Y. 2012; Krizhevsky et al. 2012) to categorise the urban frontage quality of a Google Street View image of London. Figure 8.14 shows London's predicted active frontage score derived from the deep learning model.

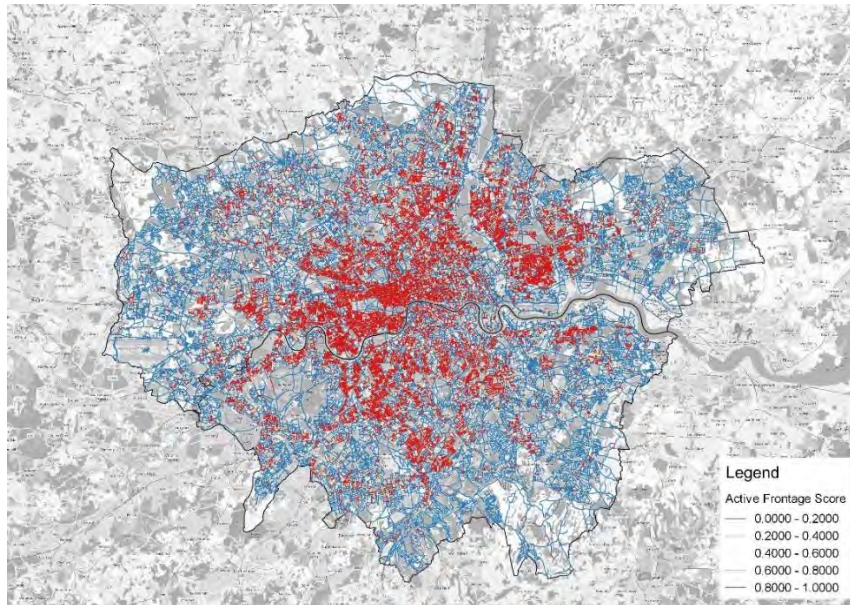


Figure 8.14 Active frontage score predicted by using a deep learning classifier. Source: Author.

This research has found encouraging results (80% accuracy) in classifying urban frontage quality using deep convolutional neural network models (Jia, Y. 2012; Krizhevsky et al. 2012). This research has also found that augmenting the baseline model with images produced by a 3D model from City-Engine can improve the overall accuracy of the model. Relating urban frontage back to house prices, early results suggests that living in a neighbourhood with greater access to high quality active frontage can also increase house prices. These findings are sensible, as street frontage quality contributes to the interest, social life, safety and success of public spaces (Heffernan 2014; Jacobs 1961). These early results point to a promising research direction in linking spatial configuration research, house prices and other urban design features of the built environment.

8.5 Research Implications

8.5.1 Implications for Planners

This section summarises the thesis by illustrating several research implications and applications. As mentioned in the beginning of the thesis, a city's fundamental advantage is its spatial configuration and the public good this configuration produces (Webster 2015). This research demonstrates that spatial configuration significantly affects a city's house prices over multiple scales. In this sense, spatial configuration can act as a lever for planners to influence the distribution of economic values in the city.

One immediate application is helping to predict the economic effects of future transport projects in London. For example, what are the economic impacts stemming from the Crossrail Project and the new connections of the Stratford Masterplan in London? What would the economic effects be if more connections were built? What are the generic accessibility effects it generates? How do these spatial accessibility effects differ between housing submarkets, and how do these new connections influence neighbourhood formation? By better understanding the positive and negative externalities of spatial configuration over multiple scales, planners are able to make more informed decisions, and thus resources can be allocated more fairly. Redistribution can then be instrumentalised through planning contribution schemes and through some forms of neighbourhood taxation. Due to the growing housing demand and rising economic inequality in London, these issues related to infrastructure planning are particularly relevant.

An advantage in using spatial network analysis as a tool in planning is these methods can be used during different stages of the urban development. From the early stages of planning such as identifying the location of development to later stages in planning in refining the design of development scheme such as the public spaces in front of stations.

8.5.2 Implications for Developers

A second implication of the research is that the results could potentially help developers identify future investment locations, reduce investment risks and improve revenue predictions. For example, by studying the errors of the hedonic price model, the developer can potentially identify locations with greater or lesser economic potential when the Crossrail is built. The research can also be used to more accurately identify neighbourhoods and housing submarkets by using community detection methods. These techniques will allow developers to better understand market segmentation, which will lead to products with stronger geographical foci. In addition, these spatial configuration techniques can also be adapted to understand and model the wider real estate market, such as the commercial sector or the industrial sectors. Most importantly, these hedonic price models can be developed into predictive models to test different master plans and to more accurately forecast future sales and prices for individual submarkets.

8.5.3 Implications for Space Syntax

Furthermore, these techniques have great implications for the field of space syntax. Importantly, this research has advanced the argument of the natural movement and the movement economy by suggesting that spatial configuration not only produces spatial network accessibility effects but also influences the construction of the community and the spatial clustering of the housing submarket. This process, in turn, impacts residential location choice and house prices. This is along the same line of enquiry as recent empirical research concerning the association between spatial configuration variables and real estate values (Chiaradia et al. 2012; Xiao et al. 2015; Law et al. 2013; 2015; 2017a).

Methodologically, this research has shown the usefulness of constructing multimodal spatial network models to consider the city as a multi-layered complex network, which can improve the correlation with socio-economic performance indicators. This research has also demonstrated the application of community detection techniques on street network dual graphs, which have the potential to be useful in identifying neighbourhood boundaries and spatial housing submarkets. Moreover, this research has exhibited how econometric techniques can be used to provide more meaningful and robust results, in addition to demonstrating how to improve interdisciplinary research.

8.6 Conclusions

This research started by asking the following question; What is the effect of spatial network configuration on intra-city house prices? Through a case study of Greater London, this research applied spatial configuration methods and the hedonic price approach in capturing the economic effects of spatial configuration across three nested levels: the property level, the local area level and the housing submarket level. The results confirm and extend the established relationship between spatial configuration measures and property values; it was further determined that the street network can be used to define local areas and housing submarkets. There are several interpretations of the positive correlation between house prices and spatial configuration properties. Concerning the significance of spatial network accessibility, one can infer buyers are valuing access to jobs, connectivity and general attractions. Regarding the importance of the local area effect, the street network may capture the urban environment more accurately, leading to stronger neighbourhood effects on house prices. Concerning housing submarket effects, street-based housing submarket might capture the spatial clustering of housing submarket more accurately leading to greater submarket effects on house prices.

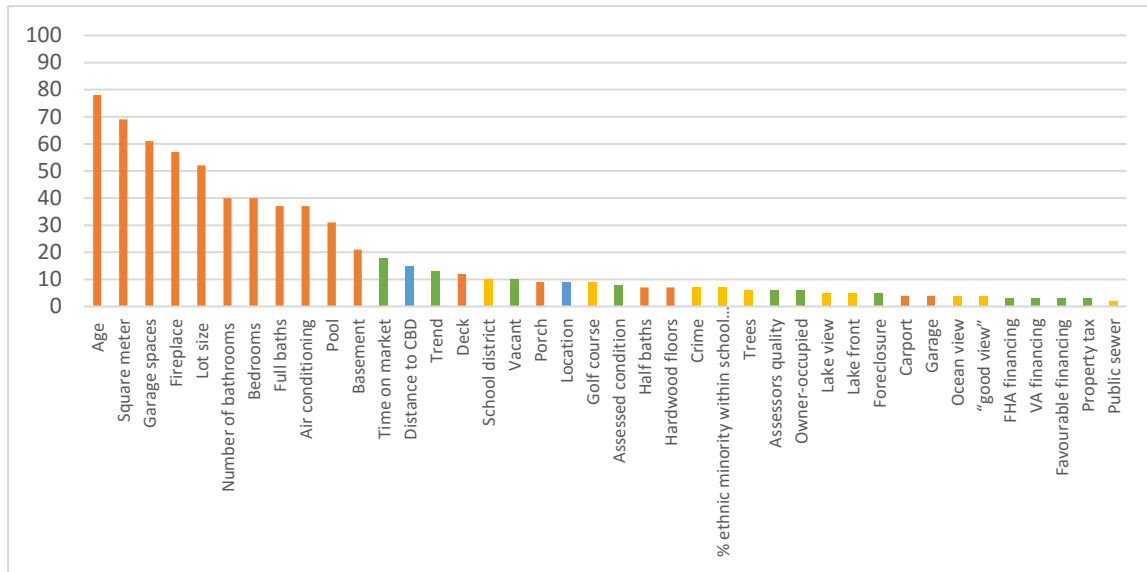
These early results are not able to infer causality but it can point to some useful directions where causal mechanism could take place. This research take the space syntax theoretical argument one step further by suggesting spatial configuration is producing the spatial network accessibility effects of centrality/employment but also influencing the construction of community and the spatial clustering of housing submarket. Spatial configuration in turn impacts upon the residential location choice and house price of a city. Importantly this research begins to spatialise the process at which house price is influenced by spatial configuration factors. As space and geometry structures the nearness and farness for people and information, understanding spatial implications is fundamental to studying cities and the markets it operates in. Extending on Glaeser's remark on "the Economic Approach of cities", where no one can make sense of cities without the tools of economics, no economist can make sense of cities without borrowing heavily from disciplines in this case; space syntax, architecture, and geography. (Glaeser 2008)

This research has several limitations and presents considerations for future research. One such limitation is that there is no clear understanding, based on the observed processes, of how spatial configuration influences the buyer's decision when purchasing a property. Future research should

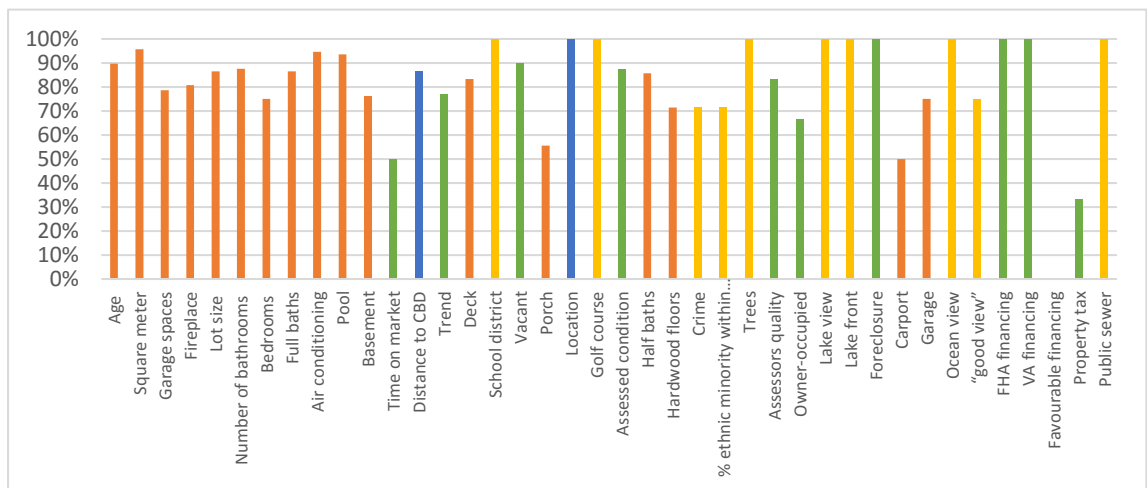
consider adopting a mixed-methods research approach, where both qualitative and quantitative methods are used, including structured interviews to observe the residential location process and advanced quantitative modelling techniques, such as agent-based-modelling or cellular automata modelling, to model these dynamic processes. With the improved availability of individual consumer behaviour data and greater computational power, it is possible to uncover these dynamic processes. Future research should continue examining spatial network configuration effects across space and time to allow for the generalisation of the results. Further research should also begin making links across all scales and developing a set of tools, theories and techniques to better understand the economic consequences of design. To end, this research contributes to a better analytical understanding of the housing market spatial configuration; this understanding improves planning decisions and redistribution, reduces developers' risks and, ultimately, builds more equitable and better-connected and more optimal neighbourhoods and cities.

Appendix A Charts made by the author based on data from Sirman et al. (2006).

Hedonic price dependent variables frequency summary.



Hedonic price dependent variables significance summary.

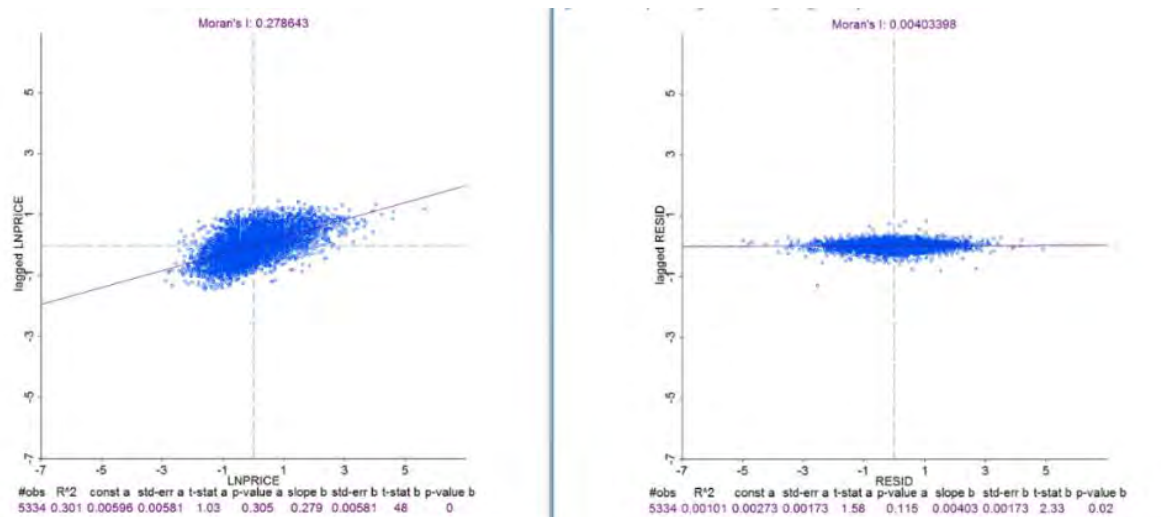


Appendix B Parks and gardens in London (English Heritage 2014).

100, CHEYNE WALK	GROSVENOR SQUARE	RICHMOND TERRACE WALK
ABNEY PARK CEMETERY	GROVE HOUSE	ROUNDWOOD PARK
ADDINGTON PALACE	GROVE PARK CEMETERY	ROYAL BOTANIC GARDENS, KEW
ALEXANDRA PALACE	GROVELANDS PARK	ROYAL HOSPITAL, CHELSEA and RANELAGH GARDENS
ARNOLD CIRCUS	GUNNERSBURY PARK	RUSKIN PARK
AVENUE HOUSE GROUNDS	HALL PLACE	RUSSELL SQUARE
BATTERSEA PARK	HAM HOUSE	SOUTH PARK GARDENS
BEDFORD SQUARE	HAMPSTEAD CEMETERY	SOUTHWARK PARK
BELAIR	HAMPTON COURT	SPRINGFIELD HOSPITAL
BENTLEY PRIORY	HAMPTON COURT HOUSE	SPRINGFIELD PARK
BERKELEY SQUARE	HANS PLACE	ST GEORGE'S GARDENS
BETHNAL GREEN GARDENS	HAREFIELD PLACE	ST JAMES'S PARK
BISHOP'S PARK	HARROW PARK	ST JAMES'S SQUARE
BLOOMSBURY SQUARE	HIGHGATE CEMETERY	ST LUKE'S GARDEN
BROCKWELL PARK	HOLLAND PARK	ST MICHAEL'S CONVENT
BROMPTON CEMETERY	HOLWOOD PARK	ST PANCRAS AND ISLINGTON CEMETERY
BROOMFIELD HOUSE	HORNIMAN GARDENS	ST PANCRAS GARDENS
BUCKINGHAM PALACE	HYDE PARK	ST PETER'S SQUARE
BUNHILL FIELDS BURIAL GROUND	INNER TEMPLE	STRAWBERRY HILL
BUSHY PARK	ISLAND GARDENS	STRAWBERRY HOUSE
CADOGAN PLACE	KENNINGTON PARK	SUNDRIDGE PARK
CANNIZARO PARK	KENSAL GREEN (ALL SOULS) CEMETERY	SYON PARK
CANONS PARK	KENSINGTON AND CHELSEA CEMETERY, HANWELL	TEDDINGTON CEMETERY
CARSHALTON HOUSE	KENSINGTON GARDENS	TERRACE AND BUCCLEUCH GARDENS
CHELSEA PHYSIC GARDEN	KENSINGTON ROOF GARDENS	THE BARBICAN
CHISWICK HOUSE	KENWOOD	THE BOLTONS
CITY OF LONDON CEMETERY	LADBROKE ESTATE	THE GROSVENOR ESTATE: BELGRAVE SQUARE
CITY OF WESTMINSTER CEMETERY	LAMBETH PALACE	THE GROSVENOR ESTATE: CHESTER SQUARE
CLISSOLD PARK	LAMORBAY PARK	THE GROSVENOR ESTATE: EATON SQUARE
COMMONWEALTH INSTITUTE	LINCOLN'S INN FIELDS	THE GROSVENOR ESTATE: WILTON CRESCENT
CORAM'S FIELDS, MECKLENBURGH and BRUNSWICK SQUARES	MANOR HOUSE GARDENS	THE HILL
CRYSTAL PALACE PARK	MARBLE HILL	THE ROOKERY

DANSON PARK	MIDDLE TEMPLE	TRAFALGAR SQUARE
DOWN HOUSE	MORDEN HALL PARK	TRENT PARK
DULWICH PARK	MYATT'S FIELDS	UPMINSTER COURT
EAST FINCHLEY CEMETERY	MYDDLETON HOUSE	VALENTINES PARK
ECCLESTON SQUARE	NORWOOD GROVE	VICTORIA EMBANKMENT GARDENS
EDWARDES SQUARE	NUNHEAD CEMETERY	VICTORIA PARK
ELTHAM PALACE	OSTERLEY PARK	VICTORIA TOWER GARDENS
FINSBURY CIRCUS	PADDINGTON CEMETERY	WALPOLE HOUSE
FINSBURY PARK	PARLIAMENT SQUARE	WALPOLE PARK
FOOTS CRAY PLACE	PECKHAM RYE PARK	WANDSWORTH PARK
FORTY HALL	POPE'S GARDEN	WANSTEAD PARK
FULHAM PALACE	PORTMAN SQUARE AND MANCHESTER SQUARE	WARWICK SQUARE
GARRICK'S VILLA	PRIMROSE HILL	WATERLOW PARK
GOLDERS GREEN CREMATORIUM	PRIORY GARDENS, ORPINGTON	WELL HALL PLEASAUNCE
GRAY'S INN	PUTNEY VALE CEMETERY	WEST HAM PARK
GREEN PARK	REGENT'S PARK	WEST NORWOOD CEMETERY AND CREMATORIUM
GREENWICH PARK	REPOSITORY WOODS	WIMBLEDON PARK
GRIMS DYKE	RICHMOND PARK	YORK HOUSE

Appendix C Global Moran's I



Appendix D Implicit prices of accessibility for individual cities in the UK (Law et al. 2017a).

Beta	2001	2011
London	0.5	0.6
Birmingham	-0.013	-0.094
Manchester	-0.09	0.005
Newcastle	0.084	0.043
Sheffield	0.029	0.04
Liverpool	-0.006	0.022
Leeds	-0.065	-0.065
Bristol	0.129	0.149
Nottingham	0.261	0.166
Bradford	-0.26	-0.044
Leicester	-0.146	-0.034
Middlesbrough	-0.013	-0.111
Portsmouth	0.122	0.016
Reading	0.113	0.038
Huddersfield	-0.14	0.024
Bournemouth	0.13	-0.056
Stoke	0.26	-0.097
Southampton	-0.064	-0.034
Preston	-0.069	-0.023
Brighton	0.38	0.38
Milton Keynes	-0.12	-0.003
Oxford	0.95	0.77
Cambridge	0.52	0.38

Bibliography

- Allen, M.T., Springer, T.M., Waller, N.G. (1995)** Implicit pricing across residential rental submarkets. *The journal of Real Estate Finance and Economics*. Volume 11, Issue 2, pp 137-151. September 1995.
- Adair, A. S., Berry, J. & McGreal, W. S. (1996)** Hedonic modelling, housing submarkets, and residential valuation, *Journal of Property Research*, 13(1), pp. 67–84.
- Adair, A.S., McGreal, S., Smyth, A., Cooper, J. Ryley, T. (2000)** "House prices and accessibility: the testing of relationships within the Belfast urban area" *Housing studies* 15 699–716.
- Addae-Dapaah, K. and Lan, Y.S. (2010).** Shopping centres and the price of proximate residential properties. 16th Annual Conference of the Pacific Rim Real Estate Society. Massey University: Wellington, New Zealand. of Economic Research, May 2005; www.nber.org/papers/w11347.pdf.
- Ahlfeldt, G. (2010)** If Alonso was right: Modeling Accessibility and Explaining the Residential Land Gradient. *Journal of Regional Science*.
- Ahlfeldt, G. (2011)** If we build, will they pay? Predicting property price effects of transport innovations. Spatial economic research centre discussion paper 75.
- Alonso, W. (1964)** *Location and Land Use: Toward a general Theory of Land Rent*. Cambridge, Massachusetts: Harvard University Press.
- Anas, A., Richard, A. , Kenneth, S.A., (1998)** Urban Spatial Structure, *Journal of Economic Literature*, American Economic Association, vol. 36(3), pages 1426-1464, September.
- Anderson, S.T., West, S.E. (2006).** Open space, residential property values, and spatial context. *Regional science and urban economics*. 36(6), pp 773-789.
- Anselin, L. (1988)** *Spatial Econometrics: Methods and Models*. Dordrecht, Kluwer.
- Anselin, L. (1995)** Local Indicators of Spatial Association- LISA. *Geographical Analysis* 27(2): 93-115.
- Armstrong, R.J. (1994)** Impacts of commuter rail service as reflected in single-family residential property values. Washington DC: Transportation Research Board (preprint, Transportation Research Board, 73rd Annual Meeting).
- Baddeley (2011)** Social Influence and Household Decision-making: a behavioural analysis of housing demand. Cambridge Working Papers in Economics. Cambridge University Library Repository. Accessed via the www: <https://www.repository.cam.ac.uk/handle/1810/241879>
- Bajic, V. (1985)** Housing-market segmentation and demand for housing attributes: some empirical findings. *AREUEA J.* 13, 58-75.
- Ball, M. and Kirwan, R. (1977)** Accessibility and supply constraints in the urban housing market" *Urban Studies* 14: 11- 32
- Banerjee, A.V. (1992)** A simple model of herd behavior. *The quarterly journal of economics*. Vol.17-3.
- Banister, D (2007)** Is Paradigm Shift too Difficult in UK Transport?, *Journal of Urban Technology* 14(2): 71-86 Online at <http://dx.doi.org/10.1080/10630730701531732>
- Bank of England (2016)** Bank of England Website. Available at: <http://www.bankofengland.co.uk/Pages/home.aspx>

- Barker, K (2004)** Review of Housing Supply. Delivering Stability: Securing our Future Housing Needs. Final Report – Recommendations, London: HMSO
- Bartik, T.J. and Smith, V.K. (1987)** Urban amenities and public policy, Handbook of Regional and Urban Economics, in: E.S.Mills (ed.), Handbook of Regional and Urban Economics, edition 1, volume 2, chapter 31, pages 1207-1254 Elsevier.
- Bates, L.K. (2006)** "Does Neighborhood Really Matter?: Comparing Historically Recognized Neighborhoods with Housing Submarkets." Journal of Planning Education and Research. 26(1):5-17.
- Batty, M. (1976)** Urban Modelling: Algorithms, Calibrations, Predictions. Cambridge, UK: Cambridge University Press.
- Batty, M. (2004)** A New Theory of Space Syntax, Working Paper 75, Centre for Advanced Spatial Analysis, UCL, London.
- Batty, M. (2009)** Notes on Accessibility, in: The Proceedings to the 7th International Space Syntax Symposium
- Batty, M. (2013)**, "The New Science of Cities" MIT Press.
- Baum-Snow, N. and Kahn, M.E. (2000)** "The effects of new public projects to expand urban rail transit," Journal of Public Economics, Elsevier, vol. 77(2): 241-263
- Bayer, P. and McMillan, R. (2005)** Sorting and Competition in Local Education Markets, NBER Working Paper No. 11802.
- Ben-Akiva, M. and Lerman, S.R. (1985).** Discrete Choice Analysis: Theory and Application to Travel Demand, The MIT Press, Cambridge.
- Benerjee, A. (1992).** A simple model of herd behaviour. The quarterly journal of economics. 107(3), 797-817.
- Benson, E. D., Hansen, J. L., Schwartz Jr., A. L., & Smersh, G. T. (1998).** Pricing residential amenities: the value of a view. The Journal of Real Estate Finance and Economics, 16(1), 55–73.
- Berger, M., Blomquist, G., and Waldner, W. (1987).** "A revealed-preference ranking of quality of life for metropolitan areas", Social Science Quarterly, 68: 761-778.
- Bettencourt, L. and West, G. (2007).** A unified theory of urban living. 2010 Oct 21;467(7318):912-3. doi:10.1038/467912a.
- Bettencourt, L. (2013).** The Origins of Scaling in Cities. Science (340) 1438-1441.
- Bettencourt, L. and Lobo (2016).** Urban Scaling in Europe. Journal of the Royal Society Interface. (13)116.
- Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., Weston, L. (2000)** Development of an Urban Accessibility Index: Literature Review. Research project conducted for the Texas Department of Transportation. Center for Transportation Research, University of Texas, Austin (TX), USA
- Black, S.E., (1999)** "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics* 114(2): 577-599.
- Blomquist, G.C., Berger, M.C. and Hoehn, J.P. (1988)** "New Estimates of Quality of Life in Urban Areas," American Economic Review, American Economic Association, vol. 78(1), pages 89-107
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre. E. (2008)** Fast unfolding of communities in

large networks. *J. Stat. Mech.*, 2008:P10008, 2008.

Boarnet, M. G. and Chalermpong, S. (2001) New Highways, House Prices, and Urban Development: A Case Study of Toll Roads on Orange County, CA, *Housing Policy Debate*, 12 (3) 575-605.

Boldi, P. & Sebastiano, V. (2014) "Axioms for centrality." *Internet Mathematics* 10.3-4 (2014): 222-262.

Bourassa, S.C., Hamelink, F., Hoesli, M., MacGregor, B.D. (1999) Defining Housing Submarkets, *Journal of Housing Economics*, 8, 2, 160-183.

Bourassa, S.C., Hoesli, M., & Sun, J. (2005) The Price of Aesthetic Externalities. *Journal of Real Estate Literature* 13 (2): 167-187.

Bourassa, S.C., Hoesli, M., Sun, J. (2004) What's in a View? *Environment and Planning A* 36 (8): 1427-1450

Bourassa, S.C., Hoesli, M., Peng, V.S. (2003) Do housing submarkets really matter? *Journal of Housing Economics*. Vol.12(1): 12-28.

Bourassa, S.C., Cantoni, E., Hoesli, M. (2007) Spatial Dependence, Housing Submarkets, and House Price Prediction. *Journal of Real Estate Finance and Economics*, 2007,35:2, 143–60.

Bowes, D.R., and Ihlanfeldt, K.R. (2001) Identifying the Impacts of Rail Transit Stations on Residential Property Values, *Journal of Urban Economics*, 50, 1-25.

Bowie, D (2010) Politics, planning and homes in a world city: the Mayor of London and strategic planning for housing in London 2000-2008, London: Routledge

Boyle, M. A., & Kiel, K. A. (2001). A survey of house price hedonic studies of the impact of environmental externalities. *Journal of real estate literature*, 9(2), 117–144. ARES. Retrieved June 23, 2011, from <http://ares.metapress.com/index/23U082061Q53QPM3.pdf>

Boyle, K.J., Poor, P.J. and Taylor, L.O. (1999) "Estimating the Demand for Protecting Freshwater Lakes from Eutrophication," *American Journal of Agricultural Economics*, Agricultural and Applied Economics Association, vol. 81(5), pages 1118-1122.

Breusch, T.S. and Pagan, A.R. (1979) A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica* 47 (5): 1287–1294.

Burgess, J., Harrison, S.M., Limb, M., (1988) People, parks and the urban green: a study of popular meanings and values for open spaces in the city. *Urban Studies* 25: 455–473.

Burgess, E.W. (1972). The Growth of the City. In *The City: Problems of Planning* (M. Steward, ed.), Penguin Books, Baltimore.

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1925). The city. Chicago, Illinois: The University of Chicago Press.

Cairncross, F. (1997). The Death of Distance: How the Communications Revolution Will Change Our Lives. Boston, MA: Harvard Business School Press.

Can, A. (1990) The measurement of neighbourhood dynamics in urban house prices, *Economic Geography*, 66, pp. 254– 272.

Carroll, T. M., Clauretje, T. M. & Jensen, J. (1996). Living next to godliness: Residential property values and churches, *Journal of Real Estate Finance and Economics*, vol. 12, pp. 319-330.

- Castells, M. (1996)**, *The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I*. Cambridge, MA; Oxford, UK: Blackwell.
- Caschili, S., De Montis, A., Chessa, A., Deplano. G. (2009)**. Weighted networks and community detection: planning productive districts in sardinia. In G. Rabino and M. Cagliani, editors, Planning, complexity and new ICT, pages 27 – 36. Alinea Editrice s.r.l., 2009
- Cervero, R., & Duncan, M. (2001)**. Rail transit's value added: Effect of proximity to light and commuter rail transit on commercial land values in Santa Clara County California. Paper prepared for National Association of Realtors Urban Land Institute.
- Chau, K.W. and Chin, T.L. (2002)** A Critical Reveiw of the Literature on the Hedonic Pricing Model and Its aplication to the Housing Market in Penang, The Seventh Asian Real Estate Society Conference. Seoul, Korea.
- Cheshire, P.C., and S. Sheppard, (1995)** 'On the Price of Land and the Value of Amenities', *Economica*, 62, 247-267.
- Chernobai, E., Reibel, M. and Carney, M. (2011)** Nonlinear Spatial and Temporal Effects of Highway Construction on House Prices, *Journal of Real Estate Finance and Economics*, 42 (3) 348-370.
- Cheshire, P. and Sheppard, S. (1998)** Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics* 60: 357–382
- Chiaradia, A. (2007)** Emergent route choice behaviour, motorway and trunk road network: the Nantes conurbation, 6th international space syntax symposium
- Chiaradia, A.J., Hillier, B., Barnes, Y. and Schwander, C. (2012a)**. Residential property value and urban form patterns in Greater London. *Urban Design and Planning, Proceedings of the Institution of Civil Engineers*
- Chiaradia, A.J., Hillier, B., Barnes, Y. and Schwander, C. (2012b)**. Compositional and urban form effects in centres in Greater London. *Urban Design and Planning: Proceedings of the Institution of Civil Engineers* 165 (1) , pp. 21-42.10.1680/udap.2012.165.1.21
- Conroy, R.C. (2001)**. The secret is to follow your nose:route path selection and angularity. *Proceedings of the 3rd International Space Syntax Symposium*. Atlanta,Georgia,USA.
- Conzen, M.R.G. (1960)** Alnwick, Northumberland: A Study in Town Plan Analysis, Institute of British Geographers, Publication no.27, London, (2nd revised edition, 1969.)
- Correll, M.R., Lillydahl, J.H. and Singell, L.D., (1978)** The Effects of Greenbelts on Residential Property Values: Some Findings on the Political Economy of Open Space, *Land Economics*, 54, issue 2, p. 207-217.
- Coulson, N.E. (1991)** Really Useful Tests of the Monocentric Model. *Land Economics*, University of Wisconsin Press, vol. 67(3), pages 299-307.
- Coulton, C. J., Korbin, J. E., Chan, T., & Su, M. (2001)**. Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology*, 29(2), 371–383.
- Court, A.T. (1939)** Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automotive Demand*, ed. Charles F. Roos 99-1 17. New York General Motors

- Croxford, B., & Penn, A. (1995).** Pedestrian exposure to urban pollution: Exploratory results. *Air Pollution III* (pp. 439-446). Southampton UK: Computational Mechanics Publications.
- Dale-Johnson, D. (1982).** "An Alternative Approach to Housing Market Segmentation Using Hedonic Pricing Data," *J. Urban Econ.* 11, 311-332.
- Dalton, R.C. and Dalton, N.S.C (2007).** "Applying Depth Decay Functions to Space Syntax Network Graphs", *Proceedings, 6th International Space Syntax Symposium, Istanbul, 2007*
- Dalton, N.S.C. (2006).** Configuration and Neighborhood: Is Place Measurable. In *Space Syntax and Spatial Cognition Workshop of the Spatial Cognition*
- Dara-Abrams, D. (2006).** Architecture of Mind and World. How Urban Form Influences Spatial Cognition. In Holscher, Conroy-Dalton, Turner (Eds). *Space Syntax and Spatial Cognition – Proceedings of the workshop held in Bremen, 24th Sept. 2006.* Bremen: Universitat Bremen, 2007.
- Davidoff, I. And Leigh, A., (2006)** How Much Do Public Schools Really Cost? Estimating the Relationship Between House Prices and School Quality, Australian National University mimeo
- Day, B. (2003).** Submarket Identification in property markets: A hedonic housing price model for Glasgow. CSERGE Working Paper Edm, No.03-09.
- Day, B, Bateman, I, Lake, I (2006)** Hedonic price analysis of road traffic noise nuisance. In: Pearce D (ed) *Environmental Valuation in Developed Countries.* Cheltenham, UK: Elgar.
- Debrezion, G., Pels, E. and Rietveld, P. (2007)** The impact of railway stations on residential and commercial property value: a meta-analysis. *Journal of real estate financial economics* 35: 161-180.
- Debrezion, G., Pels, E. and Rietveld, P. (2010)** The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market, *Urban Studies*, 48 (5)
- DeLong, B. (1990).** Liquidation Cycles: Old-Fashioned Real Business Cycle Theory and the Great Depression, NBER Working paper no. 3546.
- Department for Education (2015a)** Department for Education School Performance Database 2011 website. Available at: http://www.education.gov.uk/schools/performance/2011/download_data.html
- Department for Education (2015b)** Department for Education School Performance average score calculation. Available at: http://www.education.gov.uk/schools/performance/2013/16to18_13/average_point_score_calculations.pdf
- Des Rosiers, F., Theriault, M. (2003)** Assessing Retail Trade Areas, Local Economic Potential and Spatial Competition: How Origin-Destination Surveys May Help, *GeoSpatial Solutions*, 2003, 46–51.
- Des Rosiers, F., Theriault, M. and Ozdilek, U. (2002)** Shopping Centre Rents and Agglomeration Economies: Preliminary Findings From Empirical Evidence. Working Paper 2002-008, Faculty of Business Administration, Laval University, 2002.
- Des Rosiers, F., Lagana, A., Theriault, M., Beaudoin, M. (1996).** Shopping centres and house values : an empirical investigation. *Journal of Property Valuation & Investment.* Vol.14, Issue 4. Pp 41-62
- Desyllas, J. (1997)** Berlin in transition: analysing the relationship between land use, land value and urban morphology. *Proceedings of 5th International Space Syntax Symposium, London.*
- Deweese, D. N., (1976)** The effect of a subway on residential property values in Toronto. *Journal of*

Urban Economics, 3, 357–369.

DiPasquale, D. and Wheaton, W.C. (1994) "Housing Market Dynamics and the Future of Housing Prices," Journal of Urban Economics, 35.

Dornbusch, D. M. and Barrager, S.M. (1973) Benefit of Water Pollution Control on Property Values, report to Office of Research and Monitoring, U. S. Environmental Protection Agency, EPA-600/5-73-005

Downes, T.A. and Zabel, J.E. (2002) "The Impact of School Characteristics on House Prices: Chicago 1987-1991." Journal of Urban Economics, July 2002, 52(1): 1-25.

Dueker, K. J., & Bianco, M. J. (1999) Light rail transit impacts in Portland: The first ten years. Presented at Transportation Research Board, 78th Annual Meeting.

Edward, M. (2015) Prospects for land, rent and housing in UK cities. Future of cities: Working Paper. Government Office for Science.

Ehrenhalt, A. (2012) The great inversion and the future of American City. Knoph

Epple, D. (1987). "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." J.P.E. 95 (February): 59–80.

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. Journal of Urban Design, 14(1), 65-84.

Fack, G. and Grenet, J. (2007) Do Better Schools Raise Housing Prices? Evidence from Paris School Zoning', mimeo

Ferguson, P., Freidrich, E., & Karimi, K. (2012). Origin-destination weighting in agent modelling for pedestrian movement forecasting. Published in: Greene, M., Reyes, J. and Castro, A. eds. Proceedings: 8th International Space Syntax Symposium. Santiago de Chile: PUC, 8021:1-8021:20.

Fletcher, M., Gallimore, P., Mangan, J. (2000) Heteroscedasticity in hedonic house price models. Journal of Property Research. 17 (2), pp. 93-108.

Follain, J. R. & Jimenez, E. (1985). Estimating the demand for housing characteristics, Regional Science and Urban Economics, vol. 15, pp. 77-107.

Follain, J. R. and Malpezzi, S. (1981) Dissecting Housing Value and Rent, Washington, D.C.: Urban Institute, Working paper UI249-17

Forrest, D., Glen, J. and Ward, R. (1996) The Impact of a Light Rail System on the Structure of House Prices, Journal of Transport Economics and Policy, 30 (1) 15-29.

Fortunato, S. (2010) "Community detection in graphs". Phys. Rep. 486 (3-5): 75–174. doi:10.1016/j.physrep.2009.11.002.

Fotheringham, S., Brunsdon, C., Charlton, M. (2002) Geographically Weighted Regression: the analysis of spatially varying relationships. John Wiley & Sons, UK

Freeman, L.C. (1977) *A set of measures of centrality based on betweenness*, Sociometry 40, 35-41. pp. 37. Lehigh University.

Freeman, L. and Braconi, F. (2004) Gentrification and displacement: New York City in the 1990s. Journal of the American Planning Association 70 (1): 39-52.

Freeman, A.M. (1974) On estimating air pollution control benefits from land value studies. Journal of

Environmental Economics and Management.

Freeman, A.M. (1979) Hedonic Prices, Property Values, and Measuring Environmental Benefits: A Survey of the Issues, *Scandinavian Journal of Economics*, 81: 154-173.

Freeman, A.M. (2003) The Measurement of Environmental and Resource Values. Resources for the Future, Washington, DC.

Fujita, M. (1989) Urban Economic Theory: Land Use and City Size. Cambridge University Press, Cambridge, UK.

Galster, G. (2001). On the Nature of Neighbourhood. *Urban studies* 12:2111-2124. Publisher Full Text OpenURL

Galster, G. (1987) Residential segregation and interracial economic disparities: a simultaneous-equations approach. *Journal urban economics*. 21, 22-44.

Garrod, G.D. and Willis, K.G. (1992) Valuing Goods Characteristics: an application of the hedonic method to environmental attributes. *Journal of Environmental management*. 34(1): 59-76.

Gatzlaff, D. & Smith, M. (1993) The impact of the Miami Metrorail on the value of residences near station locations. *Land Economics*, 69 (1) 54-66. Retrieved September 21, 2005, EBSCOhost database.

Geary, R.C. (1954) The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician (The Incorporated Statistician)* 5 (3): 115–145.

Gibbons, S. and Machin, S. (2003) Valuing English Primary Schools, *Journal of Urban Economics*, 53: 197-219

Gibbons, S. (2004) The Costs of Urban Property Crime, *Economic Journal*, 114: 441-463

Gibbons, S. and Machin, S. (2005) Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1), 148-169.

Gibbons, S. and Machin, S. (2008) *Valuing school quality, better transport, and lower crime: evidence from house prices*. *Oxford review of economic policy*, 24 (1): 99-119

Gibbons, S. (2011). The amenity value of English nature: A hedonic price approach. London : SERC.

Gibbons, S. and Overman, H.G. (2012) Mostly pointless spatial econometrics. *J. Reg. Sci.* 52(2), 172–191.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin.

Girvan, M. and Newman, M.E. (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), p.7821.

Glaeser, E.L., Gottlieb, J.D., Gyourko, J. (2010) Can Cheap Credit Explain the Housing Boom? NBER Working Papers 16230, National Bureau of Economic Research, Inc.

Glaeser, E.L. and Gottlieb, J.D. (2009) "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, American Economic Association, vol. 47(4), pages 983-1028, December.

Glaeser, E.L. (2008) The Economic Approach to Cities. Faculty Research Working Paper. Harvard University, John F. Kennedy School of Government. RWP08-003.

Glaeser, E.L., Gyourko, J., Saiz, A. (2008) Housing supply and housing bubbles. NBER Working paper no. 14193.

Glaeser, E.L., Gyourko, J. (2005) "Urban Decline and Durable Housing," *Journal of Political Economy*,

University of Chicago Press, vol. 113(2), pages 345-375

Glaeser, E.L. and Ward, B.A. (2005) Myths and Realities of American Political Geography, NBER Working Papers 11857, National Bureau of Economic Research, Inc.

Glaeser, E.L. and Kahn, M. (2001) "Decentralized Employment and the Transformation of the American City," Brookings-Wharton Papers on Urban Affairs 2

Glaeser, E. (2011). Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier. New York, NY: Penguin Press.

Goldstein, H. (1987) *Multi-level Models in Educational and Social Research*. London: Charles Griffin.

Goodchild, M.F. (1986) Spatial Autocorrelation. Catmog 47, Geo Books.

Goodman A.C. (1978) 'Hedonic Price, Price Indices and Housing Markets', Journal of Urban Economics, 5, 471-484.

Goodman, A.C., Kawai, M. (1982) "Permanent Income, Hedonic Prices and Housing Demand: New Evidence," Journal of Urban Economics 12 (September 1982): 214-37.

Goodman, A.C. (1985) A note on neighbourhood size and the measurement of segregation indices. Journal of regional science, Vol.25, No.3, 1985.

Goodman, A.C. and Thibodeau, T.G. (1995) Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research, 6(1) ,25-42.

Goodman, A.C. and Thibodeau, T.G. (1998) Housing market segmentation, Journal of Housing Economics, 7(2) ,121-143.

Goodman, A.C.,Thibodeau, T.G. (2003) "Housing market segmentation and hedonic prediction accuracy" Journal of Housing Economics, Elsevier, Vol. 12(3), Pages 181-201, September.

Goodman, A.C.,Thibodeau, T.G., (2007) "The Spatial Proximity of Metropolitan Area Housing Submarkets," Real Estate Economics, American Real Estate and Urban Economics Association, vol. 35(2), pages 209-232, 06.

Greater London Authority. (2011) London geography profiles. Available at:

<http://data.london.gov.uk/dataset/ward-profiles-and-atlas> UK Open Government Licence (OGL)

Greater London Authority. (2004) London Housing Submarket Report. Greater London Authority. Available at:

http://legacy.london.gov.uk/mayor/economic_unit/docs/housing_submarkets_report_for_web.pdf

Greater London Authority (2015a) The London Plan 2011: The Spatial Development Strategy for London consolidated with alterations since 2011: March 2015. Available at:

[https://www.london.gov.uk/sites/default/files/gla_migrate_files_destination/London%20Plan%20March%202015%20\(FALP\).pdf](https://www.london.gov.uk/sites/default/files/gla_migrate_files_destination/London%20Plan%20March%202015%20(FALP).pdf)

Greater London Authority (2015b) Housing in London 2015. Available at:

https://www.london.gov.uk/sites/default/files/housing_in_london_2015.pdf

Greater London Authority (2014) Homes for London. The London Housing Strategy. Available at:

http://www.london.gov.uk/sites/default/files/Draft%20London%20Housing%20Strategy%20April%202014_0.pdf

- Griliches, Z. (1967)** "Hedonic Price Indexes Revisited: Some Notes on the State of the Art, *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, pp. 324-332
- Griliches, Z. (1971)** "Hedonic Price Indexes of Automobiles: An Econometric Analysis of Quality Change," in Zvi Griliches (ed.), *Price Indexes and Quality Change*, Cambridge: Cambridge University Press.
- Grisby, W., Baratz, M., Galster, G. and MacLenna, N.D. (1987)** *The Dynamics of Neighbourhood Change and Decline*. Oxford : Pergamon.
- Guimer`a, R., Mossa, S., Turttschi, A., Amaral. L.A.N. (2005)** The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA*, 102:7794–7799, 2005.
- Gyourko, J. and Saiz, A. (2006)** "Construction Costs And The Supply Of Housing Structure," *Journal of Regional Science*, Wiley Blackwell, vol. 46(4), pages 661-680.
- Hall, P. and Pain, K. (2006)** *Polycentric Metropolis: Learning from Mega-city Regions in Europe*. London, Sterling, VA: Earthscan.
- Handy, S. L. and Niemeier, D. A., (1997)** Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A*, 29, 1175 – 1194.
- Hansen, W. G., (1959)** How accessibility shapes land use. *Journal of the American Institute of Planners*, 25.
- Harsman, B., Quigley, J.M. (1995)** The spatial segregation of ethnic and demographic groups: comparative evidence from Stockholm and San Francisco. *Journal Urban Economics*. 37, 1-16.
- Haq, S., and Zimring, C. (2003)**. Just down the road a piece: The development of topological knowledge of building layouts. *Environment and Behavior*, 35, 132-160.
- Harris, C.D. and Ullman, E.L. (1945)**. The Nature of Cities. *Annals of the American Academy of Political and Social Sciences*, November 1945.
- Heffernan, E., Heffernan, T. & Pan, W. (2014)**. The relationship between the quality of active frontages and public perceptions of public spaces. *Urban Design international*.
- Heikkila, E., Gordon, P., Kim, J.I., Peiser, R.B., Richardson, H.W., Dale-Johnson, D., (1989)**. What happened to the CBD–distance gradient?: land values in a policentric city. *Environment and Planning* 21, 221–232.
- Helbich et al. (2013)**. Data-Driven Regionalization of Housing Markets, *Annals of the Association of American Geographers* 103, 871-889
- Helbich, M. (2015)**. "Do suburban areas impact house prices?" *Environment and Planning B: Planning and Design* 42(3) 431 – 449
- Henderson, J.V., and Mitra, A. (1996)** The New Urban Landscape: Developers and Edge Cities. *Regional Science and Urban Economics* 26 (6): 613–643
- Henderson, J. V. (1982)**. "Evaluating Consumer Amenities and Interregional Welfare Differences," *Journal of Urban Economics*, 11, 32-59.
- Henneberry, J. (1997)** Transport investments and house prices, *Journal of Property Valuation and Investment*, 16 (2) 1544-1158.

Hilber, C. and Vermeulen, W. (2010) The impacts of restricting housing supply on house prices and affordability, London: DCLG for NHPAU

<http://www.communities.gov.uk/publications/housing/impactrestrictingsupply?view=Standard>

Hildreth, P. A. (2006) Roles and Economic Potential of English Medium- Sized Cities: A Discussion Paper, Salford University of Manchester

Hillier, B. and Hanson, J. (1984). *The social logic of space*. Cambridge, MA: Cambridge University Press.

Hillier, B., Burdett, R., Peponis, J., Penn, A. (1987). "Creating Life: Or, Does Architecture Determine Anything?", *Architecture & Comportment/ Architecture & Behaviour*, 3 (3). pp. 233-250.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., Xu, J. (1993). Natural movement: or, Configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design*, 20(1), 29–66.

Hillier, B. (1996). Cities as movement economies. in: *Urban Design International* 1 (1), p. 41-60.

Hillier, B. (1999). "Centrality as a process: accounting for attraction inequalities in deformed grids", *Urban Design International* 4(3 - 4) 107-127

Hillier, B. (2002). A theory of the city as object" A Theory of the City as Object", *Urban Design International*, 7, pp.153-179.

Hillier, B. and Iida, S. (2005). Network and psychological effects in urban movement. In: Cohn, A.G. and Mark, D.M., (eds.) *Proceedings of Spatial Information Theory: International Conference, COSIT 2005, Ellicottsville, N.Y., U.S.A., September 14-18, 2005. Lecture Notes in Computer Science* (Vol. 3693). Springer-Verlag, Berlin, Germany, pp. 475-490

Hillier, B. (2006). *Space is the machine*. Cambridge, MA: Cambridge University Press.

Hillier, B. and Vaughan, L. (2007). The city as one thing. *Progress in Planning*, 67 (3), 205-230.

Hillier, B. (2009). Spatial Sustainability in Cities: Organic Patterns and Sustainable Forms. In Koch, D., Marcus, L. and Steen, J. (eds), *Proceedings 7th International Space Syntax Symposium*, Stockholm: KTH, pp. K01:1-K01:20.

Hillier, B., Yang, T., Turner, A., (2012). Normalising least angle choice in Depthmap - and how it opens up new perspectives on the global and local analysis of city space. *JOSS 2012* P155-193

Hillier, B. (2014). Spatial Analysis and Cultural Information: The need for theory as well as method in space syntax analysis. In *Spatial Analysis and Social Spaces*. Edited by Paliou, E., Undine, L., Polla, S. (2014). Walter de Gruyter GMBH, Berlin/Boston

Hillier, B. and Serra, M. (2014) Spatio-Structural Correlates at the Regional and National Scales. Presentation at the 21st International Seminar on Urban Form.

Himmelberg, C., Mayer, C., and Sinai, T. (2005) Assessing high house prices: Bubbles, fundamentals, and misperceptions. Technical report, National Bureau of Economic Research

Hoehn, J.P., Berger, M.C. and Blomquist, G.C. (1986) "A Hedonic Model of Interregional Wages, Rents and Amenity Values," Staff Papers 200902, Michigan State University, Department of Agricultural, Food, and Resource Economics.

Hough, D.E. and Kratz, C.G. (1983). Can 'Good' Architecture meet the market test? *Journal of urban*

economics 14: 40-54. 1983.

Hoyt, H. (1939). The Structure and Growth of Residential Neighborhoods in American Cities. Government Printing Office, Washington, D.C.

Hurd, R.M. (1903). Principles of City Land Values. Real Estate Record Association.

Jacobs, J. (1961). The Death and Life of Great American Cities. New York: Random House.

Jia, Y. (2012). Caffe: An open source convolution architecture for fast feature embedding.

<http://caffe.berkeleyvision.org>. Accessed: 2016.

Jiang, B., Claramunt, C., Batty, M. (1999). Geometric accessibility and geographic information: extending desktop GIS to space syntax. Computers, Environment and Urban Systems 23, 127–146.

Jones, K. (1991) Multi-level Models for Geographical Research. Norwich: Geo Books.

Jones, K. and Bullen, N. (1993) A multilevel analysis of the variations in domestic property prices: southern England 1980–1987, Urban Studies, 30, pp. 1409–1426.

Jones, K. and Bullen, N. (1994) Contextual models of urban house prices: a comparison of fixed- and random-coefficient models developed by expansion, Economic Geography, 70, pp. 252– 272.

Jones, C., Leishman, C., Watkins, C. (2003). Structural change in a local urban housing market. Environment and Planning A. 35. pp 1315-1326.

Jones, C., Leishman, C., Watkins, C. (2011) Intra-Urban Migration and Housing submarkets : theory and evidence. Housing Studies. Vo.19. Issue.2, pp 269-283.

Kain, J.F., Quigley, J.M. (1975) Housing Markets and Racial Discrimination, Columbia Press for NBER, New York, 1975.

Kain, J.F. , Quigley, J.M. (1970) Measuring the value of housing quality, Journal of the American Statistical Association 65 (440) (1970) 532–548.

Kahneman, D. (2011). Thinking, Fast and Slow, Farrar, Straus and Giroux

Kahneman, D. (2003). Maps of bounded rationality: A perspective on intuitive judgment and choice. In T. Frangsmyr (Ed.), Les Prix Nobel 2002 [Nobel Prizes 2002]. Stockholm, Sweden: Almquist & Wiksell International.

Kahneman, D., Tversky, A. (1979). "Prospect theory: An analysis of decisions under risk". Econometrica 47 (2): 313–327

Kahneman, D., Knetsch, J.L., Thaler, R.H. (1990). "Experimental tests of the endowment effect and the coase theorem". Journal of Political Economy 98, 1325–1348

Kane, T.J., Staiger, D.O. and Reigg, S.K. (1969) "School Quality, Neighborhoods and Housing Prices: The Impacts of School Desegregation." NBER Working Paper No. 11347, National Bureau

Karimi, K. (1998) Continuity and change in old cities: an analytical investigation of the spatial structure in Iranian and English historic cities before and after modernisation (Doctoral dissertation). University of London.

Karimi, K. (2000) Urban conservation and spatial transformation: preserving the fragments or maintaining the spatial spirit. URBAN DESIGN International, 5 (3-4).

Karimi, K. (2012) Special issue: Evidence-informed and analytical methods in urban design. Urban Design International, 17 (4), 253-256. doi:10.1057/udi.2012.24

Karimi, K., Parham, E., Friedrich, E., & Ferguson, P. (2013) Origin-Desitnation Weighted Choice

Model as a new tool for assessing the impact of new urban developments. Published in: Kim, Y.O., Park, H.T. and Seo, K. W. (eds.), Proceedings of the Ninth International Space Syntax Symposium, Seoul: Sejong University, Article 121

Kauko, T., Hooimeijer, P., Hakfoort, J. (2002) Capturing housing market segmentation: an alternative approach based on neural network modelling. *Housing Studies*. Vol. 17-6. P.875-894.

Kearns, A. and Parkinson, M., (2001) The significance of neighbourhood. *Urban studies* 2001, 38:2103-2110.

Khandani, A., Lo, A.W., and Merton, R.C. (2009) Systemic Risk and the Refinancing Ratchet Effect. National Bureau of Economic Research Working Paper No. 15362.

Kim, C.W., Phipps, T.T., Anselin, L. (2003) Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of Environmental Economics and Management* 45 (2003) 24-39.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing*.

Krugman, P.R. (1991) Increasing returns and economic geography. *Journal of Political Economy*, 99, 483-499.

Krugman, P., (1996). Urban concentration: the role of increasing returns and transport costs. *International Regional Science Review* 19 (1-2), 5–30.

Kwan, M.P. (2000) Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems* 2, 85–90.

Lake, I.R., Lovett, A.A., Bateman, I.J., Langford, I.H. (1998) Modelling environmental influences on property prices in an urban environment" *Computers Environment and Urban Systems* 22 121 ^ 136

Lancaster, K.J. (1971) Consumer demand: a new approach. Columbia University Press, New York.

Lancaster, K.J. (1966) A New Approach to Consumer Theory." *J.P.E.* 74 (April): 132–57.

Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: a comparative analysis *Physical Review E* 80 (5), 056117

Lancichinetti, A. and Fortunato, S. (2011). Limits of modularity maximization in community detection *Physical Review E* 84 (6), 066122

Land Registry (2014) House Price Paid Data. Data produced by Land Registry © Crown copyright 2014.

Lapham, V. (1971) "Do Blacks Pay More for Housing." *Journal of Political Economy*. 74: 132-157

Law, S., Chiaradia, Alain J. and Schwander, A., (2012). Towards a multimodal space syntax analysis: A case study of the London street and underground network. Published in: Greene, M., Reyes, J. and Castro, A. eds. Proceedings: 8th International Space Syntax Symposium. Santiago de Chile: PUC, 8021:1-8021:20.

Law, S., Karimi, K., Penn, A., Chiaradia, A. J. (2013). Measuring the influence of spatial configuration on the housing market in metropolitan London. Published in: Kim, Y.O., Park, H.T. and Seo, K. W. (eds.), Proceedings of the Ninth International Space Syntax Symposium, Seoul: Sejong University, Article 121

Law, S., Karimi, K., Penn, A. (2015). An empirical study on applying community detection methods in defining spatial housing submarkets in London. Published In: Karimi., K., Vaughan, L., Sailer, K.,

Palaiologou, G., Bolton, T. (eds.), Proceedings of the tenth space syntax symposium, London, UK., 13-17 July 2015. Space Syntax Laboratory, The Bartlett School of Architecture, University College London

Law, S., & Versluis, L., (2015). How do UK regional commuting flows relate to spatial configuration. Published In: Karimi., K., Vaughan, L., Sailer, K., Palaiologou, G., Bolton, T. (eds.), Proceedings of the tenth space syntax symposium, London, UK., 13-17 July 2015. Space Syntax Laboratory, The Bartlett School of Architecture, University College London.

Law, S., Karimi, K., Penn, A. (2017a). The Economic Value of Spatial Network Accessibility for UK Cities: A comparative analysis using the hedonic price approach. Published in the Proceedings of the Eleventh Space Syntax Symposium, Lisbon, Portugal.

Law, S., Shen, Y., Seresinhe, C. (Forthcoming). An application of convolutional neural network in street image classification. The case study of London. ACM Sigspatial workshop paper on Deep Learning.

Leamer, E.E. (2007). Housing is the Business Cycle, NBER Working Paper no. 13428.

Lebel, A., Pampalon, R., Villeneuve, P. Y. (2007). A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities in the Quebec City region. International Journal of Health Geographics 2007, 6:27 doi:10.1186/1476-072X-6-27

Leishman, C. (2009). Spatial Change and the structure of urban housing sub-markets. Housing Studies. Vol.24. Issue 5. Pp 563-585

Levinson, D. (2008) Density and Dispersion: The Co-Development of Land use and Rail in London. Journal of Economic Geography 8(1) 55-57.

Lewis-Workman, S., & Brod, D. (1997) Measuring the Neighborhood Benefits of Rail Transit Accessibility. Transportation Research Record, 1576, 147–153.

Li, M.M. and Brown, H.J. (1980) Micro-neighbourhood externalities and hedonic housing prices. Land Economics. 56: 125-140.

Linneman, P. (1981) The Demand for Residence Site Characteristics. Journal of Urban Economics 9(2): 129-148.

Linneman, P. (1980) Some Empirical results on the nature of the hedonic price function for the urban housing market. Journal of Urban Economics. 8: 47-68.

Liu, C.H., Hartzell, D.J., Greig, W., Grisson, T.V. (1990). Integration of the real estate market and the stock market: some preliminary evidence, Journal of Real Estate Finance and Economics, 3, 261-282.

Lloyd, S. P. (1957). "Least square quantization in PCM". Bell Telephone Laboratories Paper. Published in journal much later: Lloyd., S. P. (1982). "Least squares quantization in PCM". IEEE Transactions on Information Theory 28 (2): 129–137.

Löchl, M. , Axhausen, K.W. (2010) Modelling hedonic residential rents for land use and transport simulation while considering spatial effects The Journal of Transport and Land Use, Volume 3, 2010, pp. 39–63

London Borough of Barnet (NA). Hampstead Garden Suburb Conservation Area Boundary Map. Accessed via the world wide web:

http://www.barnet.gov.uk/downloads/file/2121/hampstead_garden_suburb

London Borough of Ealing (2007). Bedford Park Conservation Area Character Appraisal. Accessed via the world wide web:

[http://www2.ealing.gov.uk/ealing3/export/sites/ealingweb/services/council/committees/agendas_minut](http://www2.ealing.gov.uk/ealing3/export/sites/ealingweb/services/council/committees/agendas_minutes_reports/regulatory_committees/planning_committee/15may2007-)

[es_reports/regulatory_committees/planning_committee/15may2007-19may2008/_05_september_2007/Item_6_Bedford_Park_Conservation_Area_Report_Appendix.pdf](http://www2.ealing.gov.uk/ealing3/export/sites/ealingweb/services/council/committees/agendas_minutes_reports/regulatory_committees/planning_committee/15may2007-19may2008/_05_september_2007/Item_6_Bedford_Park_Conservation_Area_Report_Appendix.pdf)

London Chamber of Commerce and Industry (2014) Getting our house in order: The impact of housing undersupply on London-business. Available at:

<http://www.londonchamber.co.uk/DocImages/12438.pdf>

London County Council (1966). Survey of London: Volume 33 and 34. Ed F H W Sheppard.

(London 1966). Accessed via www. www.british-history.ac.uk/survey-london/vols33-34/

London Datastore (2015) London historic census population. Available at

<http://data.london.gov.uk/dataset/historic-census-population>

London Datastore (2015) London Employment data. Available at:

<http://data.london.gov.uk/dataset/gla-employment-projections>

London Datastore (2015) London average house price by Ward, LSOA, MSOA. Available at:

<http://data.london.gov.uk/dataset/average-house-prices-ward-lsoa-msoa>

London Datastore (2015) London Median Price to Earnings ratio. Available at:

<http://data.london.gov.uk/dataset/ratio-house-prices-earnings-borough>

London Datastore (2015) Statistical GIS Boundary Files for London. Available at:

<http://data.london.gov.uk/dataset/statistical-gis-boundary-files-london>

London Datastore (2015) Dwellings by Property Build Period and Type, LSOA and MSOA. Available at:

<http://data.london.gov.uk/dataset/property-build-period-lsoa>

London Datastore (2015) London housing stock by age. Available at:

<http://data.london.gov.uk/dataset/property-build-period-lsoa>

London Underground (2015). 'About London Underground', Transport for London website

<http://tube.tfl.gov.uk/content/about/default.asp>

Lowry, I. S. (1964). Model of Metropolis. Santa Monica, CA: Memorandum RM-4035-RC, Rand Corporation.

Lynch, A.K. and Rasmussen , D.W. (2001) Measuring the impact of crime on house prices. Applied Economics. 33(15): 1981-1989

Lynch, K. (1960). The image of the city, the MIT Press.

Ma, S. and Swinton, S.M. (2012) "Hedonic Valuation of Farmland Using Sale Prices versus Appraised Values," Land Economics, University of Wisconsin Press, vol. 88(1), pages 1-15.

MacLennan,D. and Tu,Y. (1996) Economic perspectives on the structure of local housing systems, Housing Studies, Vol.11, Issue 3, pp 387-406

MacQueen, J. B. (1967). "Some Methods for Classification and Analysis of Multivariate Observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 281-297.

- Marchetti, C., (1994)** Anthropological Invariants in Travel Behavior, Technological Forecasting and Social Change , 47 :75--88, Internal Publication, International Institute for Applied Systems Analysis, Laxenburg, Austria Available at: http://www.cesaremarchetti.org/archive/electronic/basic_instincts.pdf
- Marcus, L. (2000).** Architectural knowledge and urban form: The functional performance of architectural urbanity. Ph.D. thesis, KTH School of Architecture, Stockholm, Sweden.
- Marshall, A. (1890)** Principles of Economics. Macmillan, London, UK.
- Malpezzi, S., Ozanne, L., Thibodeau, T. (1980)** Characteristic Prices of Housing in 59 SMSAs. The Urban Institute.
- Malpezzi, S. (2003)** Hedonic Pricing Models: a Selective and Applied Review. In: T. O'SULLIVAN and K. GIBB, eds, Housing Economics & Public Policy. Oxford: Blackwell Publishing.
- Malpezzi, S. (2008)** Hedonic pricing models: a selective and applied review. In Housing Economics and Public Policy(O'Sullivan T and Gibb K (eds)). Blackwell Science,Oxford, UK.
- McDonald, J.F., McMillen, D.P., (1990)** "Employment subcenters and land values in a polycentric urban area: the case of Chicago,"Environment and Planning A, Pion Ltd, London, vol. 22(12), pages 1561-1574, December
- McFadden, D. (1977)** Modelling the Choice of Residential Location. Cowles Foundation Discussion Papers 477, Cowles Foundation for Research in Economics, Yale University.
- McMillen, D.P. (2012).** Quantile Regression for Spatial Data. Springer Science and Business Media.
- McMillen, D.P. (2003).** The return of centralization to Chicago: using repeat sales to identify changes in house price distance gradients, Regional Science and Urban Economics, Volume 33, Issue 3, May 2003, Pages 287-304
- McMillen, D.P. ,McDonald, J. (2004)** Reaction of house prices to a new rapid transit line: Chicago's Midway Line, 1983–1999, Real Estate Economics 32 (2004) 463–486.
- Michaels, R. and Smith, V.K. (1990)** Market segmentation and valuing amenities with hedonic models: the case of hazardous waste sites. Journal of Urban Economics 28: 223 -242.
- Mills, S.E., (1972)** Studies in the Structure of the Urban Economy. Baltimore: Johns Hopkins Press.
- Mills, S.E., (1967)** "An Aggregative Model of Resource Allocation in a Metropolitan Area," *American Economic Review* 57(2): 197-210.
- Mills, E. and Simenauer, R. (1996)** New Hedonic Estimates of Regional Constant Quality House Prices, Journal of Urban Economics, 39: 209–15.
- Molinero, C., Arcaute, E., Smith, D., Batty, M. (2015).** The Fractured Nature of British Politics. arXiv:1505.00217.
- Montello, D.R. (1991).** The measurement of cognitive distance: Methods and construct validity. Journal of Environmental Psychology 11:101-122.
- Montello, D.R. (2001).** Spatial cognition. In N. J. Smelser & P. B. Baltes (Eds.), International Encyclopedia of the Social & Behavioral Sciences. Oxford: Pergamon Press (2001): 14771-14775
- Morales, D.J., (1980)** The contribution of trees to residential property value. Journal of Arboriculture 6, 305–308.
- Moran, P.A.P. (1950)** "Notes on Continuous Stochastic Phenomena". Biometrika 37 (1): 17–23.

Moro, M., Mayor, K., Lyons, S., Tol, R. (2011). Does the housing market reflect cultural heritage? A case study of Greater Dublin. Stirling Economics Discussion Paper 2011-07

Munro, M. (1986) Testing for segmentation in the private housing market in Glasgow', DP 8, Centre for Housing Research, University of Glasgow, Glasgow

Muth, R., (1969) *Cities and Housing*. Chicago: University of Chicago Press.

Muth, R.F. and Goodman, A.C. (1989) The Economics of housing markets. A volume in the regional and urban economics section. Harwood Academic Publishers.

Nationwide (2011) London house price. produced by Nationwide.

Narvaez Zertuche, Laura. P, Penn, Alan and Griffiths, Sam. (2012). Space Syntax Economics: Decoding accessibility using property value and housing price in Cardiff, Wales. In: Greene, M and Reyes, J and Castro, A, (eds.) Eighth International Space Syntax Symposium. Pontificia Universidad Catolica de Chile: Santiago, Chile. <http://www.sss8.cl/proceedings/>

Narvaez Zertuche, L.P., Penn, A., & Griffiths, S. (2014). The Social and Economic Significance of Urban Form: A Configurative Thinking Using Bid Rent Theory. International Seminar of Urban Form and European Association for Architectural Education, 551-558. doi:10.3233/978-1-61499-365-0-558

Narvaez Zertuche, L. P. (2015). Architecture, Economy and Space: A study on the socio-economics of urban form in Cardiff, UK (Doctoral dissertation). UCL.

Nase, I., Berry, J. and Adair, A. (2013). Hedonic modelling of high street retail properties: a quality design perspective, Journal of Property Investment & Finance, Vol. 31 Issue: 2, pp.160-178.

Newman, M.E.J. (2010) Networks. An introduction. Oxford University Press.

Newman, M.E.J. and Girvan. M. (2004). Finding and evaluating community structure in networks. Phys. Rev. E, 69:026113, 2004.

Nexus Research Group (2010) London Underground and National Rail 1801-2008, University of Minnesota, Minneapolis, MN. URL: <http://nexus.umn.edu/movies.html>.

O'Sullivan, D., (2004). Complexity science and human geography. Transactions of the Institute of British Geographers 29 (3), 282-295.

Oates, W.E. (1969) The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." Journal of Political Economy 77(6): 957-71.

Openshaw, S. (1983). The modifiable areal unit problem. Norwick: Geo Books.

Office National Statistics (2015) Office for National Statistics Geography. Available at: <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/electoral-wards-divisions/index.html>

Office for National Statistics (2015) Census Geography products for England and Wales. Open Geography Portal. Accessed via <https://geoportal.statistics.gov.uk/geoportal/catalog/main/home.page>

Ordnance Survey (2014). Ordnance Survey Open Data Meridian 2 Dataset. Open source. Access on 01/08/2014 from the world wide web <http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html>. © Crown Copyright {2014}

Ordnance Survey (2015). Ordnance Survey Open Data Boundary Line Dataset. Open source. Access on 18/03/2015 from the world wide web <http://www.ordnancesurvey.co.uk/business->

and-government/products/.opendata-products.html. © Crown Copyright {2015}

Orford, S. (1999). Valuing the Built Environment: GIS and House Price Analysis. Aldershot: Ashgate.

Orford, S. (2001) Modelling spatial structures in local housing market dynamics: a multilevel perspective, *Urban Studies*, 37(9), pp. 1643–1671.

Orman, G.K., Labatut, V., Cherifi, H. (2012). Comparative evaluation of community detection algorithms: A topological approach. *Journal of Statistical Mechanics* 8, P08,001 (2012)

Ortiz, C., Hillier, B. (2005). In search of patterns of Land-use in Mexico City using Logistic regression at the plot level. 6th International Space Syntax Symposium

Osland, L., Thorsen, I., Gitlesen, J.P. (2007) Housing Price Gradients in a Geography with one Dominating Center. *Journal of Real Estate Research*, 2007, 29:3, 321–346

Pace, R.K. and Otis, W.G. (1997) Using the Spatial Configuration of the Data to Improve Estimation, *Journal of Real Estate Finance and Economics* 14(3): 333-340.

Pace, R.K., Barry, R., Clapp, J.M., Rodriguez, M. (1998) Spatiotemporal Autoregressive Models of Neighbourhood Effects, *Journal of Real Estate Finance and Economics*. 17(1):15-33.

Palm, R. (1978). Spatial segmentation of the urban housing market. *Economic Geography* 54. 210-221

Palm, R. (1976) The role of real estate agents as information mediators in two american cities. *Geografiska Annaler. Series B, Human Geography*. 58(1): 28-41.

Palmquist, R. B. (1984) "Estimating the demand for the characteristics of housing", *Review of Economics and Statistics*, 66, pp 394-404.

Park, J. (2011) Spatial analysis of housing markets using land rent theory of political economy: the cases of London, Seoul and Los Angeles, PhD thesis, UCL Bartlett

Park, J. (2013) The division of spatial housing submarkets : a theory and the case of Seoul. *Environment and Planning A*. Vol.45 pp 668-690

Penn, A., & Croxford, B. (1997). Effects of street grid configuration on kerbside concentrations of vehicular emissions. *Proceedings of First International Symposium on Space Syntax*, 16-18 April, University College London, London.

Penn, A., Hillier, B., Banister, D., Xu, J. (1998a). Configurational modelling of urban movement networks, *Environment and Planning B: Planning and Design*, 24, p. 59-84.

Penn, A., & Croxford, B. (1998b). Fingerprinting Urban Kerbside Carbon Monoxide Concentrations: interaction between street grid configuration, vehicle flows and local wind effects. *International Journal of Vehicle Design*, 20 (1-4 (Special Issue)), 60-70. doi:10.1504/IJVD.1998.001835

Penn, A., Desyllas, J., Vaughan, L. (1999). "The space of innovation: interaction and communication in the work environment" *Environment and Planning B: Planning and Design* 26(2) 193 – 218.

Penn, A., Turner, A. (2002). Space syntax based agent models. In Schreckenberg, M., Sharma, S., eds.: *Pedestrian and Evacuation Dynamics*. Springer-Verlag, Heidelberg, Germany (2002) 99–114.

Penn, A. (2003). Space syntax and spatial cognition: or why the axial line? *Environment and Behavior* 35 (2003) 30–65.

Pennington, G., Topham, N., Ward, R. (1990) Aircraft Noise and Residential Property Values Adjacent

to Manchester International Airport. *Journal of Transport Economics and Policy*, 24: 49-59.

Peponis, J. (1988) Social structure of space and relations of community in six Greek cities, Secretariat of Research and Technology, Athens, Ministry of Energy Industry, Research and technology.

Peponis, J., Wineman, J., Rashid, M., Kim, S.H., Bafna, S., (1997). "On the description of shape and spatial configuration inside buildings: convex partitions and their local properties", *Environment and Planning (B): Planning and Design*, 24, 761-781.

Peponis, J., Zimring, C., Choi-Yoon, K. (1990). "Finding the Building in Wayfinding". *Environment and Behavior*, v.22, n.5, 555-590.

Piketty, T. (2013) *Capital in the Twenty-First Century*, London

Pons, P., Latapy, M. (2005). "Computing communities in large networks using random walks", *Computer and Information Sciences - Iscis 2005, Proceedings*, vol. 3733, pp. 284-293, 2005.

Pons, P., Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications* Vol.10-2. P.191-218.

Porta, S., Crucitti, P., Latora, V. (2006), The network analysis of urban streets: a dual approach *Physica A: Statistical Mechanics and its Applications* 369 (2), 853-866.

Puga, D., (2010). The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science* 50 (1), 203–219.

Quan, D.C. and Titman, S. (1999). Do real estate price and stock prices move together? An international analysis? *Real Estate Economics*, 27(2), 183-207.

Quercia, D., Schifanella, R., Aiello, L.M. (2014) The shortest path to happiness: Recommending beautiful, quiet and happy routes in the city. In *Proceedings of ACM HT*.

Quercia, D., Schifanella, R., Aiello, L.M., Mclean, K. (2015) Smelly Maps: the digital life of urban smellscape. In *ICWSM'15: Proceedings of the 9th AAAI International Conference on Weblogs and Social Media*. AAAI.

Raghavan, U. N., Albert, R., Kumara, S. (2007) *Phys. Rev.E* 76(3), 036106.

Rasmussen, S.E. (1937) *London: The Unique City*. The Macmillan Company. New York.

Ratti, C. (2004). Space syntax: some inconsistencies. *Environment and Planning B: Planning and Design* 2004, v. 31, p. 487-499

Read, S. (1999). Space syntax and the Dutch city. *Environment and Planning B*, 26, pp.251–264

Reichardt, J. and Bornholdt, S. (2004). "Statistical mechanics of community detection", *Phys. Rev. E*, vol. 74, no. 1, p. 016110, 2006.

Ridker, R.G. and Henning, J.A. (1967) 'The Determinants of Residential Property Values With Special Reference to Air Pollution', *Review of Economics and Statistics* 49, 246--257.

Roback, J. (1982) Wages, rents, and the quality of life. *Journal of Political Economy*, 90(1), 257–78.

Rodriguez, M. and Sirmans, C. F. (1994) Quantifying the Value of a View in Single-Family Housing Markets. *Appraisal Journal*, 62, 600-603.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34-55.

Rossum, G. (2007) Python programming language. In *USENIX Annual Technical Conference*.

Rosvall, M., Bergstrom, C. (2007). "An information-theoretic framework for resolving community

structure in complex networks", Proceedings of the National Academy of Sciences of the United States of America, vol. 104, no. 18, pp. 7327-7331, 2007.

Rosvall, M., Bergstrom, C. (2008). "Maps of random walks on complex networks reveal community structure", Proceedings of the National Academy of Sciences, vol. 105, no. 4, p. 1118, 2008.

Royal Parks (2014) The Royal Parks Website. Accessed via the world wide web. <https://www.royalparks.org.uk/>

Sabidussi, G. (1966) The centrality index of a graph. Psychometrika 31, 581–603.

Sampson, R.J. (2012) Greater American City: Chicago and the enduring neighbourhood effect. University of Chicago Press.

Schafer, R. (1979). Racial discrimination in the Boston housing market Journal of Urban Economics, vol. 6, pp. 176-196.

Schelling, T. (1969). Models of Segregation. The American Economic Review, 59(2): 488-493.

Scheurer, J., Curtis, C. (2007). Accessibility measures: Overview and practical applications. Urbanet Working Paper No.4 march 2007.

Schnare, A and Struyk, R. (1976) Segmentation in urban housing markets, Journal of Urban Economics, 3, 146-166.

Schnare, A.B. (1980) Trends in residential segregation by race: 1960-1970 Journal Urban Economics. 7. 293-301.

Scoppa, M., Peponis, J., (2015) "Distributed attraction: the effects of street network connectivity upon the distribution of retail frontage in the City of Buenos Aires" Environment and Planning B: Planning & Design 42 354-378

Schwander, C. (2007) NETWORK ANALYSIS APPLIED: the railway network in South East England, 6th International Space Syntax Symposium

Sheppard, S. (1999) Hedonic analysis of housing markets in "Handbook of Regional and Urban Economics," Elsevier Edition 1, volume 3, number 3

Sheppard, F.H.W. (1966) 'Map of the Parish of St. Anne, Soho', in Survey of London: Volumes 33 and 34, St Anne Soho, ed. F H W Sheppard (London, 1966), accessed via the world wide web. <http://www.british-history.ac.uk/survey-london/vols33-4/map-of-st-anne-soho> [accessed 12 January 2015].

Shelter and KPMG (2014) Building the homes we need: a programme for the 2015 government, London. Available at: www.shelter.org.uk/__data/assets/pdf_file/0019/802270/Building_the_homes_we_need_-_a_programme_for_the_2015_government.pdf

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. Cities, 55, 9–21.

Shiller, R.J. (2000) Irrational Exuberance. Princeton University Press.

Shiller, R.J. (2008) The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It, Princeton University Press.

Simon, H. (1957) A behavioral model of rational choice. In Models of man, social and rational: Mathematical essays on rational human behaviour in a social setting. New York: Wiley

- Sirmans, C. F. and Guidry, K.A. (1993)** The determinants of shopping centre rents, *Journal of Real Estate Research* 8(1): 107–15.
- Sirmans, G.S., MacDonald, L., MacPherson, D.A., Zietz, E.N. (2006)** The Value of Housing Characteristics: A Meta Analysis. *Journal of Real Estate Finance and Economics*, Vol. 33, No. 3.
- Sirpal, R. (1994).** Empirical modeling of the relative impacts of various sizes of shopping centres on the value of surrounding residential properties, *Journal of Real Estate Research*, vol. 9, no. 4, pp. 487-505.
- Smith, V.K. and Huang, J.C. (1995)** Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. *Journal of Political Economy*, 103: 209-27.
- Smith, D. (2010)** Valuing Housing and Green Spaces: Understanding Local Amenities, the Built Environment and House Prices in London. GLA Economics, London, Working Paper 42.
- Smith, D (2011)** Polycentricity and Sustainable Urban Form: An Intra-Urban Study of Accessibility, Employment and Travel Sustainability for the Strategic Planning of the London Region, PhD, UCL
- Sobolevsky, S., Campari, R., Belyi, A., & Ratti, C. (2014).** A general optimization technique for high quality community detection in complex networks. *Phys. Rev. E* 90 (1), 012811,
- Sonstelie, J.C. and Portney, P.R. (1980)** Gross rents and market values: testing the implications of Tiebout's hypothesis. *Journal of Urban Economics* 7 :102 -118
- Stata (2012).** Multilevel and longitudinal modelling using Stata, Third edition. Stata Press.
- Strasheim, M. (1975)** An Econometric Analysis of the Urban Housing market. Urban and Regional Studies Number 2. National Bureau of Economic Research. New York. Columbia University Press.
- Thamesmead Trust (2007)** Thamesmead Street Map produced by Gallions Housing Association. Available at: <http://www.trust-thamesmead.co.uk/>
- Theriault, M. Des Rosiers, F. and Joerin, F. (2005).** Modelling accessibility to urban services using fuzzy logic. *Journal of property investment and finance*. 23(10) pp. 22.
- Tobler, W., (1970)** "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(2): 234-240.
- Tounonen, A. and Law, S. (2017).** Affiliation spatially expressed: How Social Networks structure Residential Mobilities in London. Published in the Proceedings of the Eleventh Space Syntax Symposium, Lisbon, Portugal.
- Tiebout, C. (1956)** "A Pure Theory of Local Expenditures" *Journal of Political Economy* 64(5): 416-424.
- Transport for London (2015)** London Travel demand survey data. Available at: <https://tfl.gov.uk/corporate/publications-and-reports/london-travel-demand-survey>
- Tse, R.Y.C.(2002)** Estimating Neighbourhood Effects in House Prices: Towards a New Hedonic Model Approach. *Urban Studies* 39(7): 1165-1180.
- Turner, A. (2007)** "From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis" *Environment and Planning B: Planning and Design* 34(3) 539 – 555
- Turner, A. (2000).** Angular analysis: a method for the quantification of space, Working Paper 23, Centre for Advanced Spatial Analysis, UCL, UK. pp.3;

- Ullman, E.L., Harris, C.D. (1945).** "The Nature of Cities". *The Annals of the American Academy of Political and Social Science*. 242: 7–17.
- Vandell, K.D. and Lane, J.S. (1989).** The economics of architecture and urban design: some preliminary findings. AREUEA journal. Vol.17-2.
- Varoudis, T. (2012)** depthmapX multi-platform spatial network analysis software. Version 0.30 Open source. Available at: <http://varoudis.github.io/depthmapX/>
- Vaughan, L. (2005).** The relationship between physical segregation and social marginalization in the urban environment. *World Architecture*, No.185, pp. 88-96.
- Von Thunen, J. (1826)** *Isolated State: An English Edition of Der isolierte Staat* (translated by Wartenberg C and Hall P). Pergamon Press, Oxford, UK.
- Voith, R. (1993)** Changing capitalization of CBD-oriented transportation systems: Evidence from Philadelphia, 1970–1988. *Journal of Urban Economics*, 33, 361–376.
- Walter, T. (1878),** 'Soho', in *Old and New London: Volume 3* (London, 1878), pp. 173-184 Originally published by Cassell, Petter & Galpin, London, 1878. Accessed via the world wide web: <http://www.british-history.ac.uk/old-new-london/vol3/pp173-184> [accessed 5 January 2015].
- Watkins, C.A. (2001)** "The definition and identification of housing submarkets" *Environment and Planning A* 33(12) 2235 – 2253
- Waddell, P. (2002)** UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning, *Journal of the American Planning Association*, 68(3), pp. 297-314.
- Walter, F. and Schlapfer, F. (2010)** Landscape amenities and local development: A review of migration, regional and hedoni pricing studies. *Ecological Economics* 70: 141-152.
- Webster C, (2010).** "Pricing accessibility: urban morphology, design and missing markets" *Progress in Planning* 73 77–111
- Wegener, M. (1994)** Operational urban models: state of the art, *Journal of the American Planning Association*, 60, pp. 17–29.
- Wegener, M. (2004).** Overview of land-use transport models. *Handbook of Transport Geography and Spatial Systems* 5: 127–146.
- Whitehand, J.W.R. (1981)** *The Urban Landscape: Historical Development and Management*. Papers by M. R. G. Conzen., Institute of British Geographers Special Publication no.13, Academic Press, London.
- Wikipedia. (2015).** Soho. Wikipedia. Accessed via the world wide web: http://wikitravel.org/en/File:Areas_of_Central_London_I.png
- Wilson, A.G. (1970)** *Entropy in urban and regional modelling: Monographs in spatial and environmental systems analysis* 1. Pion, London.
- Wilson, A. (2000).** *Complex Spatial Systems: The modelling foundations of urban and regional analysis*. Pearson Education.
- Xiao, Y., Webster, C., Orford, S. (2015)** Urban Configuration, Accessibility, and Property Prices: A Case study of Cardiff, Wales. *Environment and Planning B: Planning and Design*. 2015. Vol.42, Pages 1-22.

- Xiao, Y., Lu, Yi., Guo, Y., Yuan, Y. (2017).** Estimating the willingness to pay for green space services in Shanghai: Implications for social equity in urban china. *Urban Forestry and Urban Greening*. 26.95-103.
- Yang, T. and Hillier, B. (2007)** The fuzzy boundary: the spatial definition of urban areas. In *Proceedings 6th international space syntax symposium*. Istanbul, Turkey, ed. A.S.Kubat, Ö Ertekin, and Y.lo. Güney, 091.01-22. Cenkler, Istanbul: Istanbul Technical University.
- Zabel, J. (2008).** Using hedonic models to measure racial discrimination and prejudice in the U.S. Housing market. *Hedonic methods in housing markets*. Springer. pp.177-201.